Joint Angular- and Delay-Domain MIMO Propagation Parameter Estimation Using Approximate ML Method

Cássio B. Ribeiro, *Member, IEEE*, Andreas Richter, *Member, IEEE*, and Visa Koivunen, *Member, IEEE*

***Abstract*—In this paper, we derive an estimation method that jointly estimates the parameters of the concentrated propagation paths and the distributed scattering component that are frequently observed in multiple-input multiple-output (MIMO) channel sounding measurements. The joint angular-delay domain model leads to a correlation matrix with high dimensionality, which makes direct implementation of a maximum-likelihood (ML) estimator unfeasible. We derive low-complexity methods for computing approximate ML estimates that exploit the structure of the covariance matrices. We propose an iterative two-step procedure that alternates between the estimation of the parameters of the concentrated propagation paths and the parameters of the distributed scattering. For the distributed scattering, the estimator first optimizes the parameters describing their time-delay structure. Then, using the estimated time-delay parameters, the parameters of the angular distributions are optimized. We present simulation results and compare the estimated time-delay and angular distributions to the actual distributions, demonstrating that high-quality estimates are obtained. The large sample performance of the estimator is studied by establishing the Cramér–Rao lower bound (CRLB) and comparing it to the variances of the estimates. The simulations show that the variance of the proposed estimation technique reaches the CRLB for relatively small sample size for most parameters, and no bias is observed.**

***Index Terms*—Channel parameter estimation, maximum-likelihood estimation, multipath channels, radio propagation.**

NOMENCLATURE

$\mathbf{b}(\varphi_{\mathrm{R},\mathrm{l}}, \varphi_{\mathrm{T},\mathrm{l}})$	Array response to $\varphi_{\mathrm{R},\mathrm{l}}$ and $\varphi_{\mathrm{T},\mathrm{l}}$.
$b_{m_1 m_2}$	$b_{m_1 m_2} = j2\pi d_{m_1 m_2}$.
$\mathbf{C}_h$	$E[\mathbf{h}\mathbf{h}^H]$.
$\mathbf{C}_y$	$M_o \times M_o$ covariance matrix of the received signal.
$\mathbf{C}_{\bar{y}}$	Covariance matrix of $\bar{\mathcal{Y}}_m$.
$\mathbf{C}_{\bar{y}}$	Covariance matrix of $\bar{\mathcal{Y}}_m$.
$\mathbf{C}_w$	$E[\mathbf{w}\mathbf{w}^H]$.
$d_{m_1 m_2}$	Distance between elements m_1 and m_2 in the receive array.
$\mathbf{F}$	Normalized DFT matrix.
$\mathbf{h}$	Spatial content of the diffuse scattering component (DSC).
$\mathcal{I}$	Fisher information matrix.
M_f	Number of frequency samples.
M_o	$M_o = M_r M_t M_f$.
M_r	Number of RX antennas.
M_s	Number of observations.
M_t	Number of TX antennas.
N_{sp}	Number of specular propagation paths.
P	Number of clusters.
$\bar{\mathbf{s}}$	One observation of the specular component.
$\mathbf{v}(\boldsymbol{\theta}_w)$	Sampled version of the PDP in frequency domain.
$\mathbf{V}_y, \mathbf{V}_w, \mathbf{V}_h$	Eigenvectors of $\mathbf{C}_y, \mathbf{C}_w, \mathbf{C}_h$, respectively.
$\mathbf{w}$	One observation of the frequency content of the DSC.
$\mathbf{y}(f)$	Received signal in frequency domain.
α_1	Maximum power.
β_d	Normalized coherence bandwidth of the DSC.
ϵ_p	Mixture proportions of the DSC.
$\varphi_{\mathrm{R},\mathrm{l}}$	Receive azimuth angle of the lth specular path.
$\varphi_{\mathrm{T},\mathrm{l}}$	Transmit azimuth angle of the lth specular path.
γ_l	Complex gain of specular component.
κ_p	Scattering about the symmetry center.
$\boldsymbol{\Lambda}_y, \boldsymbol{\Lambda}_w, \boldsymbol{\Lambda}_h$	Eigenvalues of $\mathbf{C}_y, \mathbf{C}_w, \mathbf{C}_h$, respectively.
μ_p	Symmetry center or "mean angle."
$\boldsymbol{\theta}_w$	Frequency-domain DSC parameters: $\boldsymbol{\theta}_w = \{\alpha_1, \beta_d, \tau_d\}$.
$\boldsymbol{\Theta}_h$	Angular-domain DSC parameters: $\boldsymbol{\Theta}_h = \{\mu_p, \kappa_p, \epsilon_p\}$, $p = 1, \ldots, P$.
$\boldsymbol{\theta}_{h,p}$	Angular-domain DSC parameters for cluster p: $\boldsymbol{\theta}_{h,p} = \{\mu_p, \kappa_p, \epsilon_p\}$.
$\boldsymbol{\theta}_{\mathrm{wn}}$	Frequency-domain DSC parameters plus noise variance.
$\boldsymbol{\theta}_{\mathrm{dn}}$	DSC parameters plus noise variance $\{\boldsymbol{\theta}_{\mathrm{wn}}, \boldsymbol{\theta}_{\mathrm{h}}\}$.

Manuscript received March 31, 2006; revised December 8, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yonina C. Eldar.

C. B. Ribeiro and A. Richter are with the Helsinki University of Technology, Signal Processing Laboratory, FIN-02015, Finland (e-mail: cassio.ribeiro@hut.fi; arichter@wooster.hut.fi).

V. Koivunen is with Helsinki University of Technology, SMARAD CoE, Signal Processing Lab, FIN-02015, Finland (e-mail: visa.koivunen@hut.fi).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2007.896247

I. INTRODUCTION

THE interest in the multidimensional structure of the mobile radio channel is growing rapidly. This is mainly due to the fact that future beyond 3G wireless systems will employ multiantenna transceivers in order to improve spectral efficiency and radio link quality. Consequently, realistic channel models that are verified by real-world measurement campaigns are

needed especially for transceiver design and network planning purposes. Channel sounding and related propagation parameter estimation are key tasks in creating such channel models. In particular, the double-directional modeling of the radio channel has attracted a lot of interest because it gives a better physical insight into the wave propagation mechanism in real-world radio environments [1], and it has the ability to remove the measurement antenna influence from the channel observation. Moreover, developing and comparing the performance of various multiple-input multiple-output (MIMO) transceiver structures requires such advanced channel models.

In channel parameter estimation from channel sounding measurements, two types of models have mainly been employed so far. A widely used model approximates the observed radio channel with a large number of discrete (deterministic) waves (see, e.g., [2] and [3]). The other approach describes the radio channel as a zero-mean circular complex Gaussian process [4], [5]. The former model is well suited to describe dominant concentrated (specular) propagation paths, i.e., paths with a small delay and angular spread, each contributing significantly to the wave propagation between transmitter (TX) and receiver (RX). The latter model is well suited to describe distributed (diffuse) scattering (DDS). It has been shown in [6] that both propagation mechanisms contribute significantly to the wave propagation. In [3], a data model for channel parameter estimation is proposed, which combines these two models. It is shown that an estimator that accounts for both concentrated propagation paths and distributed scattering outperforms estimators that ignore either of the channel components. However, in [3], the estimator is derived assuming that the contribution of the distributed scattering is an independent and identically distributed (i.i.d.) process in the angular domain at TX and RX. It only accounts for the correlation of the distributed scattering in the time delay and the frequency domain and ignores the correlation in angular domain. So far, no channel parameter estimation results have been published that provide more information about the angular properties of DDS in radio channels. This is mainly due to the fact that no estimator has been developed that estimates also the spatial correlation of the DDS. The model and the algorithm derived in this paper fill this gap.

A maximum-likelihood estimator for scattered sources is proposed in [7], where the angular distribution is assumed to be Gaussian. The angular spread is assumed to be small so that the correlation between the antenna elements can be approximated by a Taylor series expansion. In [8], for small angular spreads, the authors propose a first-order Taylor expansion of the spatial signature of each source. This leads to the generalized array manifold (GAM) model, which can be used in conjunction with well-known subspace-based methods, such as MUSIC. This method provides a parametric model for an instantaneous realization of the fading channel, as shown in [9]. Again, by assuming small angular spreads, the authors in [10] show that it is possible to approximate a scattered source as a combination of two rays symmetrically located around the nominal direction. This approximation allows using computationally efficient algorithms such as ESPRIT and root-MUSIC. The resulting algorithms are called Spread ESPRIT, Spread root-MUSIC, and so on [10]. Other methods stemming from these can be found, e.g., in [11] and [12]. These methods are derived only for narrowband channels.

In this paper, we derive an approximate ML estimator for channel parameter estimation from channel sounding measurements, which also takes the angular distribution of the distributed scattering component into account. It estimates the parameters of the concentrated propagation paths and the distributed scattering component jointly. The power-delay profile of the scattering component is modeled using an exponential distribution, which is typically observed in measurement campaigns [1]. The angular power profile is modeled using a mixture of angular von Mises distributions. The mixture distributions are employed in order to estimate the propagation parameters in scattering environments with multiple clusters of scatterers with high fidelity [13]. The von Mises distribution is described in detail in [14].

We propose a two-step procedure that alternates between the estimation of the parameters of the concentrated propagation paths and the parameters of the distributed scattering. For the distributed scattering component, the time-delay domain parameters modeled using an exponential distribution are estimated first. The angular domain parameters are then improved by fixing the time-delay domain parameters. This way, the correlation of the distributed scattering in the time-delay and frequency domains is fully taken into account. A whitening transformation is applied to the data prior to the estimation of the specular component [15], thus decoupling the problem of estimating the specular and diffuse components. As a result, well-known estimation methods can be used to estimate the parameters of the specular components with low complexity, like the SAGE-based procedure in [2] and [16].

The joint angular-delay model for the distributed scattering leads to a correlation matrix with high dimensionality. This makes the use of a maximum-likelihood (ML) estimator using finite precision arithmetics and finite memory resources unfeasible. In this paper, computationally efficient methods for finding approximate ML estimates are derived. The structure of the covariance matrices is fully exploited in the process.

Moreover, we derive the Cramér–Rao lower bound (CRLB) for the problem. It gives the lower bound on the variance by any unbiased estimator. We show in simulation that the variance of the estimates is close to the bound for a relatively small number of iterations and small number of samples.

This paper is organized as follows: in Section II, we describe the channel model considered in this paper. In Section III, the alternating technique for parameter estimation is described. In Section IV, the Cramér–Rao lower bound is established. Finally, in Section V, we present some simulation results and compare the large sample performance of the estimation technique to the CRLB.

II. Signal Model

Assuming a channel sounding arrangement with M_r antennas at the receiver and M_t antennas at the transmitter, the signal at the receiver in frequency-domain is given by

$$\mathbf{y}(f) = \mathbf{s}(f) + \mathbf{n}_d(f) + \mathbf{n}(f) \tag{1}$$

where $\mathbf{s}(f)$ represents the specular components of the propagation paths, $\mathbf{n}_d(f)$ represents the diffuse scattering component, and $\mathbf{n}(f)$ represents the zero-mean complex Gaussian measurement noise. The sampled data are arranged in vector $\mathbf{y}(f)$ as

$$\mathbf{y}(f) = \mathrm{vec}\left(\begin{bmatrix} y_{1,1}(f) & \cdots & y_{1,M_t}(f) \\ \vdots & \ddots & \vdots \\ y_{M_r,1}(f) & \cdots & y_{M_r,M_t}(f) \end{bmatrix}\right)$$

where the vec operator stacks the columns of a matrix. The specular components are modeled as

$$\begin{aligned}\mathbf{s}(f) &= \sum_{l=1}^{N_{\mathrm{sp}}} \mathbf{s}_l(f) \\ &= \sum_{l=1}^{N_{\mathrm{sp}}} \gamma_l \mathbf{b}(\varphi_{\mathrm{R},l}, \varphi_{\mathrm{T},l}) \exp(-j2\pi f \tau_l) u(f) \end{aligned} \tag{2}$$

where $u(f)$ is the transmitted signal, γ_l is the complex gain, $\mathbf{b}(\varphi_{\mathrm{R},l}, \varphi_{\mathrm{T},l})$ is the array response to receive azimuth angle $\varphi_{\mathrm{R},l}$ and transmit azimuth angle $\varphi_{\mathrm{T},l}$, and τ_l is the normalized delay for path $l = 1, \ldots, N_{\mathrm{sp}}$. The array response is given as a function of the receive and transmit array responses, $\mathbf{b}(\varphi_{\mathrm{R},l})$ and $\mathbf{b}(\varphi_{\mathrm{T},l})$, respectively, as $\mathbf{b}(\varphi_{\mathrm{R},l}, \varphi_{\mathrm{T},l}) = \mathbf{b}(\varphi_{\mathrm{R},l}) \otimes \mathbf{b}(\varphi_{\mathrm{T},l})$, where $\otimes$ denotes the Kronecker product.

We assume that the excitation signal $u(f)$ is a multicarrier spread-spectrum signal (MCSSS) [3], which is designed such that $|u(f)|$ is constant over the bandwidth of interest. Therefore, a raw estimate of the channel observation can be determined without changing the statistics of the noise $\mathbf{n}(f)$, simply by dividing the received samples by the known excitation function $u(f)$. Hence, (1) can be rewritten as

$$\mathbf{y}(f) = \mathbf{s}(f) + \mathbf{n}_d(f) + \mathbf{n}(f). \tag{3}$$

Let M_f be the number of observed frequency samples. We then define the $M_o \times 1$ vector $\mathcal{Y}$ as

$$\begin{aligned}\mathcal{Y} &= \begin{bmatrix} \mathbf{y}(0) \\ \vdots \\ \mathbf{y}(M_f - 1) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{s}(0) + \mathbf{n}_d(0) + \mathbf{n}(0) \\ \vdots \\ \mathbf{s}(M_f - 1) + \mathbf{n}_d(M_f - 1) + \mathbf{n}(M_f - 1) \end{bmatrix} \\ &= \bar{\mathbf{s}} + \bar{\mathbf{n}}_d + \bar{\mathbf{n}} \end{aligned} \tag{4}$$

where $M_o = M_r M_t M_f$, and

$$\bar{\mathbf{s}} = [\mathbf{s}^{\mathrm{T}}(0) \quad \cdots \quad \mathbf{s}^{\mathrm{T}}(M_f - 1)]^{\mathrm{T}} \tag{5}$$

$$\bar{\mathbf{n}}_d = [\mathbf{n}_d^{\mathrm{T}}(0) \quad \cdots \quad \mathbf{n}_d^{\mathrm{T}}(M_f - 1)]^{\mathrm{T}} \tag{6}$$

$$\bar{\mathbf{n}} = [\mathbf{n}^{\mathrm{T}}(0) \quad \cdots \quad \mathbf{n}^{\mathrm{T}}(M_f - 1)]^{\mathrm{T}}. \tag{7}$$

The channel sounding technique assumed in this paper is based on time-division multiplexing of each transmitter and receiver antenna. Such technique is used, for example, in the PropSound channel sounder used in [16]. This particular structure makes it possible to measure the channel between each transmit and receive antenna pair independently. For further discussion of radio channel sounding techniques, see [17]. Deterministic maximum-likelihood estimation techniques such as the SAGE-based method in [16] represent the received signal as a combination of a large number of discrete waves. Consequently, parameters from each wave must be estimated. Hence, the algorithms often have convergence problems, and the estimates contain artifacts due to local maxima in the likelihood function and high dimensionality of the parameter space.

The following assumptions are employed throughout this article:

a) the process $\mathbf{n}_d(f)$ is zero-mean complex temporally white circular Gaussian;
b) the channel can be treated as constant during the time it takes to measure one realization of the channel;
c) the additive noise $\bar{\mathbf{n}}$ is an i.i.d. zero-mean circular complex Gaussian process with known covariance matrix, $\mathbf{C}_n = E[\bar{\mathbf{n}}\bar{\mathbf{n}}^H] = \sigma_n^2\mathbf{I}$, and independent of $\bar{\mathbf{n}}_d$ and $\bar{\mathbf{s}}$.

Assumption a) comes from the observation that a very large number of waves, having independent weights, reach the receiver from scattered sources. Therefore, the central limit theorem can be used in this case to show that the received data are Gaussian distributed. The covariance matrix of $\mathbf{n}_d(f)$ is obtained by writing the DSC as a sum of a very large number (N) of specular-like components, each one modeled similarly to (2). Hence

$$\mathbf{n}_d = \frac{1}{\sqrt{NP_a}} \sum_{i=1}^{N} \alpha_i \mathbf{w}_i \otimes \mathbf{h}_i \tag{8}$$

where α_i is the complex gain, $P_a = E[|\alpha|^2]$, the vector $\mathbf{h}_i$ of dimension $M_r M_t \times 1$ represents the spatial content of the DSC and is a function of the array response, and the vector $\mathbf{w}_i$ of dimension $M_f \times 1$ represents the frequency-dependent content of the DSC. Assume also that α_i, $\mathbf{w}_i$, and $\mathbf{h}_i$ are zero-mean and independent $\forall i$. The covariance matrix of $\mathbf{n}_d$ is given by

$$\begin{aligned} &E\left[\mathbf{n}_d\mathbf{n}_d^H\right] \\ &\quad = \frac{1}{NP_a} E\left[\left(\sum_{i=1}^{N} \alpha_i \mathbf{w}_i \otimes \mathbf{h}_i\right)\left(\sum_{j=1}^{N} \alpha_j^* \mathbf{w}_j^H \otimes \mathbf{h}_j^H\right)\right] \\ &\quad = \frac{1}{NP_a} \sum_{i=1}^{N} E\left[\left|\alpha_i^2\right| \left(\mathbf{w}_i\mathbf{w}_i^H\right) \otimes \left(\mathbf{h}_i\mathbf{h}_i^H\right)\right] \\ &\quad = \frac{1}{NP_a} \sum_{i=1}^{N} P_a E\left[\mathbf{w}_i\mathbf{w}_i^H\right] \otimes E\left[\mathbf{h}_i\mathbf{h}_i^H\right] \\ &\quad = E[\mathbf{w}\mathbf{w}^H] \otimes E[\mathbf{h}\mathbf{h}^H]. \end{aligned} \tag{9}$$

Based on the assumptions above, the PDF of the received signal $\mathcal{Y}$ is completely characterized by its mean, $E[\mathcal{Y}] = \bar{\mathbf{s}}$, and its $M_o \times M_o$ covariance matrix

$$\begin{aligned}\mathbf{C}_y &= E[(\mathcal{Y} - \bar{\mathbf{s}})(\mathcal{Y} - \bar{\mathbf{s}})^H] = E\left[(\bar{\mathbf{n}}_d + \bar{\mathbf{n}})\left(\bar{\mathbf{n}}_d^H + \bar{\mathbf{n}}\right)^H\right] \\ &= E[\mathbf{w}\mathbf{w}^H] \otimes E[\mathbf{h}\mathbf{h}^H] + E[\bar{\mathbf{n}}\bar{\mathbf{n}}^H] \\ &= \mathbf{C}_w \otimes \mathbf{C}_h + \sigma_n^2\mathbf{I} \end{aligned} \tag{10}$$

where $\mathbf{I}$ is the $M_o \times M_o$ identity matrix, $\mathbf{C}_w = E[\mathbf{w}\mathbf{w}^H]$, and $\mathbf{C}_h = E[\mathbf{h}\mathbf{h}^H]$. Observe that our model is based on the assumption that the covariance matrix of the DSC can be factorized into

a Kronecker product (10). It is not assumed that the covariance matrix of the complete MIMO channel observation can be factorized into a Kronecker product.

A. Delay and Frequency-Domain Characterization

For the delay domain, we use the model in [3], which is based on the observation that the power delay profile (PDP) has an exponential decay over time and a base delay that is related to the distance between the transmitter and receiver. The PDP for infinite bandwidth is given by

$$\psi(\tau) = E[|\mathsf{w}(\tau)|^2] = \begin{cases} 0, & \tau < \tau_d' \\ \alpha_1/2, & \tau = \tau_d' \\ \alpha_1 e^{-B_d(\tau-\tau_d')}, & \tau > \tau_d' \end{cases} \quad (11)$$

where B_d is the coherence bandwidth, α_1 denotes the maximum power, and τ_d' is the base delay.

The Fourier transform of (11), the correlation function of the channel in the frequency domain, is given by

$$\psi(\Delta f) = \frac{\alpha_1}{\beta_d + j2\pi\Delta f} e^{-j2\pi\Delta f\tau_d'} \quad (12)$$

where $\beta_d = B_d/B_m$ is the normalized coherence bandwidth, and B_m is the measurement bandwidth. Let us define the sampled version of the correlation function $\mathbf{v}(\boldsymbol{\theta}_w)$, $\boldsymbol{\theta}_w = \{\alpha_1, \beta_d, \tau_d\}$, in frequency domain as

$$\mathbf{v}(\boldsymbol{\theta}_w) = \frac{\alpha_1}{M_f}\left[\frac{1}{\beta_d} \quad \frac{e^{-j2\pi\tau_d}}{\beta_d + j\frac{2\pi}{M_f}} \quad \cdots \quad \frac{e^{-j2\pi(M_f-1)\tau_d}}{\beta_d + j2\pi\frac{M_f-1}{M_f}}\right] \quad (13)$$

where τ_d is the normalized base delay.

Since the process is stationary, the correlation between components at different frequencies is given by

$$\Psi(f_1, f_2) = \psi(f_1 - f_2). \quad (14)$$

Hence, the covariance matrix of the diffuse scattering (assuming the received signal is spatially white) may be modeled as a Toeplitz matrix

$$\mathbf{C}_w = \mathrm{toep}(\mathbf{v}(\boldsymbol{\theta}_w), \mathbf{v}(\boldsymbol{\theta}_w)^H) \quad (15)$$

where $\mathrm{toep}(\mathbf{a}, \mathbf{b}^H)$ denotes a Toeplitz matrix with $\mathbf{a}$ as its first column and $\mathbf{b^H}$ as its first row, with $a_1 = b_1^*$.

B. Angular-Domain Characterization

Using the extended Saleh–Valenzuela (SVA) channel model in [4], we can write $E[\mathbf{h}\mathbf{h}^H]$ as a function of the angular parameters. A similar model is obtained in [5] and [18] following a different approach that is related to the geometry of the distribution of scatterers. Assuming that for the diffuse scattering there is statistical independence between AOAs and AODs, we can use the modeling strategy in [4] to characterize the covariance matrix of the channel in angular domain as

$$\mathbf{C}_h = \mathbf{C}_h^{\mathrm{R}} \otimes \mathbf{C}_h^{\mathrm{T}} \quad (16)$$

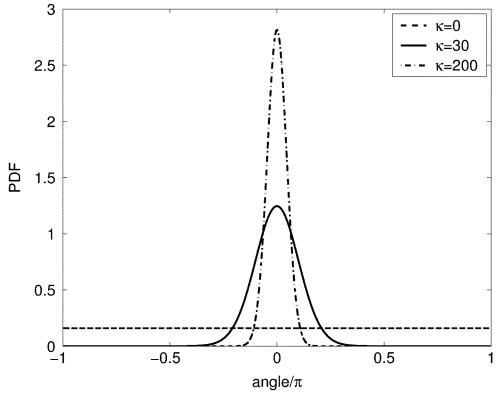

Fig. 1. von Mises PDF for different values of κ, with $\mu = 0$.

where $\mathbf{C}_h^{\mathrm{R}}$ and $\mathbf{C}_h^{\mathrm{T}}$ denote the covariance matrix at the receiver and transmitter side, respectively. From a parameter estimation point of view, the extension from the spatially uncorrelated to the correlated case is more complex than extension from single input multiple output (SIMO) to MIMO, as long as the Kronecker structure in (16) holds. Hence, we will limit the discussion to unidirectional estimation, but the results can be naturally extended to the double directional case. We also assume for simplicity that an uniform linear array (ULA) is used at the receiver. Then the correlation at the receiver side is given by

$$\mathbf{C}_{h,m_1m_2}(\boldsymbol{\Theta}_h) = \int_{-\pi}^{\pi} \exp(b_{m_1m_2}\cos(\phi)) f(\phi, \boldsymbol{\Theta}_h)\, d\phi \quad (17)$$

where $f(\phi, \boldsymbol{\Theta}_h)$ is any angular PDF of ϕ, characterized by parameters $\boldsymbol{\Theta}_h$, $b_{m_1m_2} = j2\pi d_{m_1m_2}$, and $d_{m_1m_2}$ is the distance between elements m_1 and m_2 in the receive array. An angular PDF must at least satisfy $f(\phi, \boldsymbol{\Theta}_h) = f(\phi + 2\pi k, \boldsymbol{\Theta}_h)\ \forall\, k \in \mathbb{Z}$, with $\phi \in [\phi_0, \phi_0 + 2\pi)$, $\phi_0 \in \mathbb{R}$. Hence, a Gaussian PDF that has an infinite support is not suitable. The von Mises distribution [14] defined in angular domain is more appropriate. It is defined as follows:

$$f(\phi, \boldsymbol{\Theta}_h) = \frac{1}{2\pi I_0(\kappa)} \exp(\kappa\cos(\phi - \mu)) \quad (18)$$

where μ is the symmetry center or "mean angle" and $I_0(\cdot)$ is the modified Bessel function of the first kind of order zero. The parameter κ is related to the variance of the von Mises distribution, i.e., how scattered about the symmetry center the data are. It can be chosen between 0 (isotropic scattering) and ∞ (extremely concentrated). Fig. 1 illustrates the von Mises PDF for different values of κ.

In channel measurements, multimodal angular PDFs are often found as a result of signals arriving from a number of different clusters. This can be modeled using a mixture of angular PDFs

$$f(\phi, \boldsymbol{\Theta}_h) = \sum_{p=1}^{P} \epsilon_p f_p(\phi, \boldsymbol{\theta}_{h,p}) \quad (19)$$

where P is the number of mixture components, $\sum_{p=1}^{P} \epsilon_p = 1$, ϵ_p are unknown mixture proportions, and $f_p(\phi)$ is any valid angular PDF. With this definition of the angular PDF, the angular domain parameters are defined as $\boldsymbol{\Theta}_h = \{\boldsymbol{\theta}_{h,1}, \ldots, \boldsymbol{\theta}_{h,P}\}$, with $\boldsymbol{\theta}_{h,p} = \{\mu_p, \kappa_p, \epsilon_p\}$, $p = 1, \ldots, P$.

Using (18), the cross correlation in (17) may be written analytically as [5], [18]

$$\mathbf{C}_{h,m_1m_2}(\boldsymbol{\Theta}_h) = \sum_{p=1}^{P} \epsilon_p \frac{I_0\left(\left\{\kappa_p^2 + b_{m_1m_2}^2 + 2\kappa_p b_{m_1m_2}\cos(\mu_l)\right\}^{\frac{1}{2}}\right)}{I_0(\kappa_p)}. \quad (20)$$

III. Parameter Estimation

Let us denote by $\mathcal{Y}_m$ the mth observation of $\mathcal{Y}$, $m = 1, \ldots, M_s$. Assuming $\mathcal{Y}$ is circular complex Gaussian and that the realizations $\mathcal{Y}_m$ are i.i.d., we can write the log-likelihood function as

$$L(\mathcal{Y}_1, \ldots, \mathcal{Y}_{M_s}) = -M_o M_s \log \pi - M_s \log |\mathbf{C}_y| - \sum_{m=1}^{M_s} (\mathcal{Y}_m - \bar{\mathbf{s}})^H \mathbf{C}_y^{-1} (\mathcal{Y}_m - \bar{\mathbf{s}}) \quad (21)$$

where M_s is the number of observations. We will also assume that the noise is circular complex white Gaussian with variance σ_n^2.

A method that estimates $\mathbf{C}_w$ and the noise variance is proposed in [3] assuming the input signal to be spatially white, i.e., $\mathbf{C}_h = \mathbf{I}$. The method exploits the Toeplitz structure of $\mathbf{C}_w$ for the computation of the ML estimates. This reduces the computational complexity by avoiding the direct computation of determinants and matrix inversions. Unfortunately, it is not possible to directly extend the method for joint estimation of $\mathbf{C}_w$, $\mathbf{C}_h$, and σ_n^2, since $\mathbf{C}_y$ in (10) is not Toeplitz in general. Also, direct optimization of the likelihood function using (10) is not feasible due to the high dimensionality of the matrices involved. In current sounding systems, typical values for M_f and M_r, M_t are in the range $M_f = [100, 2000]$, and $M_r, M_t = [4 \ldots 64]$. But with the rapid development of the channel sounders, these values may grow. This leads to a dimension of $\mathbf{C}_y$ ranging from 400×400 to $128\,000 \times 128\,000$, or even higher.

We propose an estimation method that reduces the computational complexity by using the following iterative procedure.

1) Optimize for the parameters of the specular components such as azimuth and elevation angle of arrival/departure, time delay, doppler spread etc., using the previously estimated covariance matrix.
2) Remove the contribution of the specular components from data and optimize for the covariance matrix of the diffuse scattering components plus noise variance.
3) Repeat the procedure until convergence or a maximum number of iterations is reached.

Step 2) can be further decomposed into two steps:

2a) Optimize for the frequency-domain parameters and noise variance.
2b) Optimize for the angular-domain parameters, with $\mathbf{C}_w$ as calculated in the previous step.

This iterative procedure is illustrated in Fig. 2.

A key benefit of the proposed method is the separate optimization of specular and diffuse scattering components. This reduces the number of variables for each local optimization. A similar approach has been considered in [3], where the signal was assumed to be spatially white.

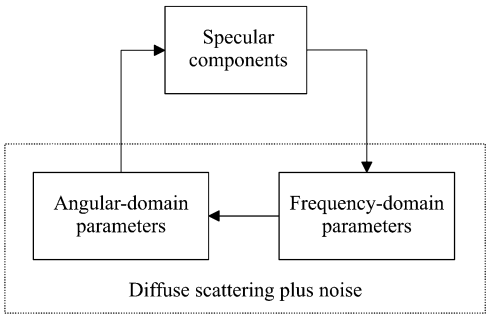

Fig. 2. Two-step procedure for joint optimization of specular components and diffuse scattering parameters.

The further decomposition of Step 2) into Steps 2a) and 2b) is important due to the high dimensionality of the matrices involved. With this two-step procedure, it is possible to exploit the Toeplitz structure of $\mathbf{C}_w$ for the computation of the ML estimates. Also, the covariance matrix manipulated in Step 2b) is only $\mathbf{C}_h$, which is typically much smaller in dimension than $\mathbf{C}_w$.

A. Specular Component

Most algorithms for parameter estimation of specular components assume that any additive noiselike components are white. Hence, they cannot be directly applied to the model used in this paper, unless the covariance matrix of $\mathbf{n}_d(f) + \mathbf{n}(f)$ is the identity matrix. In order to avoid this problem and allow the use of well-known low-complexity algorithms for parameter estimation of specular components, a *prewhitening* transform is applied to the data such that its covariance matrix becomes a constant times the identity matrix [15]. We define the prewhitening matrix $\mathbf{U}$, such that

$$E[(\mathbf{U}^{-H}\mathcal{Y} - \mathbf{U}^{-H}\bar{\mathbf{s}})(\mathbf{U}^{-H}\mathcal{Y} - \mathbf{U}^{-H}\bar{\mathbf{s}})^H] = \mathbf{U}^{-H}\left(\mathbf{C}_w \otimes \mathbf{C}_h + \sigma_n^2\mathbf{I}\right)\mathbf{U}^{-1} = \mathbf{I}. \quad (22)$$

Therefore, $\mathbf{U}^{-H}\mathcal{Y}$ can be used to estimate the parameters of the specular-like propagation paths using any well-known algorithm, such as the SAGE-based procedure in [2] and [16].

The matrix $\mathbf{U}$ is any square root matrix of $\mathbf{C}_y$ such that $\mathbf{C}_y = \mathbf{U}^H\mathbf{U}$. It can be obtained, e.g., by the Cholesky decomposition of $\mathbf{C}_y$, since the presence of additive noise guarantees that $\mathbf{C}_y$ is positive definite. Another possibility to calculate $\mathbf{U}$ is through the eigenvalue decomposition of $\mathbf{C}_y$. This approach proves to be useful in the next sections. Let $\mathbf{V}_y$ and $\boldsymbol{\Lambda}_y$ denote the eigenvectors and eigenvalues of $\mathbf{C}_y$, i.e.,

$$\mathbf{C}_y = \mathbf{V}_y\boldsymbol{\Lambda}_y\mathbf{V}_y^H \quad (23)$$

where we have used the fact that the eigenvectors of a Hermitian matrix are unitary. Similarly, we can also define $\mathbf{V}_w$, $\boldsymbol{\Lambda}_w$, $\mathbf{V}_h$, and $\boldsymbol{\Lambda}_h$, such that $\mathbf{C}_w = \mathbf{V}_w\boldsymbol{\Lambda}_w\mathbf{V}_w^H$ and $\mathbf{C}_h = \mathbf{V}_h\boldsymbol{\Lambda}_h\mathbf{V}_h^H$. We can then write

$$\mathbf{C}_y = \left(\mathbf{V}_w\boldsymbol{\Lambda}_w\mathbf{V}_w^H\right) \otimes \left(\mathbf{V}_h\boldsymbol{\Lambda}_h\mathbf{V}_h^H\right) + \sigma_n^2\mathbf{I} = \left(\mathbf{V}_w \otimes \mathbf{V}_h\right)\left(\boldsymbol{\Lambda}_w \otimes \boldsymbol{\Lambda}_h + \sigma_n^2\mathbf{I}\right)\left(\mathbf{V}_w^H \otimes \mathbf{V}_h^H\right). \quad (24)$$

Comparing (24) with (23), we conclude that

$$\mathbf{V}_y = \mathbf{V}_w \otimes \mathbf{V}_h \tag{25}$$

$$\mathbf{\Lambda}_y = \mathbf{\Lambda}_w \otimes \mathbf{\Lambda}_h + \sigma_n^2 \mathbf{I}. \tag{26}$$

Hence, $\mathbf{U}^{-H}$ can be defined as

$$\mathbf{U}^{-H} = \left(\mathbf{\Lambda}_w \otimes \mathbf{\Lambda}_h + \sigma_n^2 \mathbf{I}\right)^{-1/2} \left(\mathbf{V}_w^H \otimes \mathbf{V}_h^H\right). \tag{27}$$

The product $\mathbf{U}^{-H}\mathcal{Y}$ can be computed efficiently as

$$\mathbf{U}^{-H}\mathcal{Y} = \operatorname{diag}\left(\left(\mathbf{\Lambda}_w \otimes \mathbf{\Lambda}_h + \sigma_n^2 \mathbf{I}\right)^{-1/2}\right) \circ \operatorname{vec}\left[\left(\mathbf{V}_w^H \left(\mathbf{V}_h^H \operatorname{mat}(\mathcal{Y}_m, M_r M_t, M_f)^{\mathrm{T}}\right)^{\mathrm{T}}\right] \tag{28}$$

where the mat operator reshapes a vector into a matrix, as follows:

$$\operatorname{mat}\left(\begin{bmatrix} \mathbf{a}_1^{m\times 1} \\ \vdots \\ \mathbf{a}_N^{m\times 1} \end{bmatrix}, M, N\right) = \begin{bmatrix} \mathbf{a}_1^{m\times 1} & \cdots & \mathbf{a}_N^{m\times 1} \end{bmatrix} \tag{29}$$

the vec operator stacks the columns of a matrix, and $\circ$ denotes elementwise multiplication, i.e., the Hadamard product.

The benefits of this implementation over the Cholesky decomposition is that the eigenvalues and eigenvectors of $\mathbf{C}_w$ and $\mathbf{C}_h$ can be used to simplify the estimator of the diffuse scattering component. Also, it will be shown in the next section that the computation of $\mathbf{V}_w^H$ and its multiplication by a vector can be done in a very efficient way. The Kronecker product in (28) has to be computed only for the diagonal elements, and the multiplication of a diagonal matrix and a vector has a complexity of $\mathcal{O}(M_o)$.

B. Frequency-Domain Parameters

In the sequel, we will assume that the specular components have been estimated and removed from the observation, and hence the likelihood function is given by

$$L(\mathcal{Y}_1, \ldots, \mathcal{Y}_{M_s}) = -M_o M_s \log \pi - M_s \log |\mathbf{C}_y| - \sum_{m=1}^{M_s} \mathcal{Y}_m^H \mathbf{C}_y^{-1} \mathcal{Y}_m. \tag{30}$$

An estimator for the frequency-domain parameters is derived in [3]. It is assumed that the channel covariance matrix has the structure $\mathbf{C}_y = (\mathbf{C}_w + \sigma_n^2 \mathbf{I}) \otimes \mathbf{I}$, i.e., the channel is assumed to be spatially white. This is a special case of the situation expressed in (20), with $P = 1$ and $\kappa = 0$. Since we do not assume whiteness in the spatial domain, the method in [3] is not directly applicable here.

Using the decomposition of the estimation procedure outlined in Section III already allows some simplification in the optimization. This is due to lower dimensional searches. However, we still have not solved the problem of calculating the determinant and inverse of $\mathbf{C}_y$ at every iteration. These computations can be simplified by writing $\mathbf{C}_y$ as a function of its eigenvalues and eigenvectors, as in Section III-A.

We can exploit the Kronecker structure of the eigenvalues and eigenvectors of $\mathbf{C}_y$ in order to simplify the optimization procedure. This allows us to compute only the eigenvalues and eigenvectors of $\mathbf{C}_w$ and $\mathbf{C}_h$, and then obtain $\mathbf{V}_y$ and $\mathbf{\Lambda}_y$. The logarithm of the determinant of $\mathbf{C}_y$ can now be calculated as

$$\begin{aligned} \log |\mathbf{C}_y| &= \log\left(|\mathbf{V}_w \otimes \mathbf{V}_h| \left|\mathbf{\Lambda}_w \otimes \mathbf{\Lambda}_h + \sigma_n^2 \mathbf{I}\right| \left|\mathbf{V}_w^H \otimes \mathbf{V}_h^H\right|\right) \\ &= \log \prod_{j=1}^{M_o} \left(\boldsymbol{\lambda}_w \otimes \boldsymbol{\lambda}_h + \sigma_n^2 \mathbf{1}_{M_o}\right)_{\{j\}} \\ &= \sum_{j=1}^{M_o} \log\left(\left(\boldsymbol{\lambda}_w \otimes \boldsymbol{\lambda}_h + \sigma_n^2 \mathbf{1}_{M_o}\right)_{\{j\}}\right) \end{aligned} \tag{31}$$

where $\boldsymbol{\lambda}_w$ and $\boldsymbol{\lambda}_h$ are vectors containing the eigenvalues of $\mathbf{C}_w$ and $\mathbf{C}_h$, respectively, $\mathbf{1}_{M_o}$ is a $M_o \times 1$ vector whose entries are equal to 1, and $(\cdot)_{\{j\}}$ denotes the jth element of $(\cdot)$. It is clear that the computational complexity of calculating the determinant is reduced. Another important observation is that the exchange of the order in which the log is computed allows for easier implementation with finite precision. This is due to the fact that the eigenvalues $\boldsymbol{\lambda}_w$ can have a large spread, since they are an approximation to the PDP in (11).

The computation of $\mathbf{C}_y^{-1}$ can also be simplified using

$$\begin{aligned} \mathbf{C}_y^{-1} &= \left[\left(\mathbf{V}_w \otimes \mathbf{V}_h\right)\left(\mathbf{\Lambda}_w \otimes \mathbf{\Lambda}_h + \sigma_n^2 \mathbf{I}\right)\left(\mathbf{V}_w^H \otimes \mathbf{V}_h^H\right)\right]^{-1} \\ &= \left(\mathbf{V}_w \otimes \mathbf{V}_h\right)\left(\mathbf{\Lambda}_w \otimes \mathbf{\Lambda}_h + \sigma_n^2 \mathbf{I}\right)^{-1}\left(\mathbf{V}_w^H \otimes \mathbf{V}_h^H\right). \end{aligned} \tag{32}$$

Further simplifications are possible if we take into account that $\mathbf{C}_h$ is fixed while optimizing for $\mathbf{C}_w$. Let us rewrite the likelihood function in (30) as a function of $\mathbf{V}_w$, $\mathbf{\Lambda}_w$, $\mathbf{V}_h$, $\mathbf{\Lambda}_h$, and σ_n^2

$$\begin{aligned} L(\mathcal{Y}_1, \ldots, \mathcal{Y}_{M_s}) \propto &- \sum_{j=1}^{M_r M_t M_f} \log\left(\left(\boldsymbol{\lambda}_w \otimes \boldsymbol{\lambda}_h + \sigma_n^2 \mathbf{1}\right)_{\{j\}}\right) \\ &- \frac{1}{M_s} \sum_{m=1}^{M_s} \mathcal{Y}_m^H (\mathbf{V}_w \otimes \mathbf{V}_h) \\ &\times \left(\mathbf{\Lambda}_w \otimes \mathbf{\Lambda}_h + \sigma_n^2 \mathbf{I}\right)^{-1} \left(\mathbf{V}_w^H \otimes \mathbf{V}_h^H\right) \mathcal{Y}_m. \end{aligned} \tag{33}$$

Clearly, it is not necessary to calculate $\mathbf{V}_w^H \otimes \mathbf{V}_h^H$ and the multiplication of this $M_o \times M_o$ matrix by $\mathcal{Y}_m$ at every iteration of (30), since $\mathbf{V}_h$ and $\mathcal{Y}_m$ are fixed. In order to simplify the problem, we will define the transformed signal

$$\bar{\mathcal{Y}}_m = \left(\mathbf{I}_{M_f} \otimes \mathbf{\Lambda}_h^{-1/2} \mathbf{V}_h^H\right) \mathcal{Y}_m \tag{34}$$

where it is assumed that $\mathbf{C}_h$ is nonsingular. An efficient form for computation of $\bar{\mathcal{Y}}_m$ is given by

$$\bar{\mathcal{Y}}_m = \operatorname{vec}\left(\mathbf{\Lambda}_h^{-1/2} \mathbf{V}_h^H \operatorname{mat}(\bar{\mathcal{Y}}_m, M_r M_t, M_f)\right). \tag{35}$$

The covariance matrix of $\bar{\mathcal{Y}}_m$ is given by

$$\begin{aligned} \mathbf{C}_{\bar{y}} &= \left(\mathbf{I}_{M_f} \otimes \mathbf{\Lambda}_h^{-1/2} \mathbf{V}_h^H\right) E\left[\mathcal{Y}_m \mathcal{Y}_m^H\right] \left(\mathbf{I}_{M_f} \otimes \mathbf{V}_h \mathbf{\Lambda}_h^{-1/2}\right) \\ &= \left(\mathbf{I}_{M_f} \otimes \mathbf{\Lambda}_h^{-1/2} \mathbf{V}_h^H\right) \left(\mathbf{V}_w \otimes \mathbf{V}_h\right) \left(\mathbf{\Lambda}_w \otimes \mathbf{\Lambda}_h + \sigma_n^2 \mathbf{I}\right) \\ &\quad \times \left(\mathbf{V}_w^H \otimes \mathbf{V}_h^H\right) \left(\mathbf{I}_{M_f} \otimes \mathbf{V}_h \mathbf{\Lambda}_h^{-1/2}\right) \\ &= \left(\mathbf{V}_w \otimes \mathbf{I}_{M_r M_t}\right) \left(\mathbf{\Lambda}_w \otimes \mathbf{I}_{M_r M_t} + \sigma_n^2 \mathbf{I}_{M_f} \otimes \mathbf{\Lambda}_h^{-1}\right) \\ &\quad \times \left(\mathbf{V}_w^H \otimes \mathbf{I}_{M_r M_t}\right). \end{aligned} \tag{36}$$

The assumption that $\mathbf{C}_h$ is nonsingular implies that $\mathbf{C}_{\bar{y}}$ is positive definite. Even though the covariance matrix $\mathbf{C}_{\bar{y}}$ is not

block diagonal, it is clear that the elements of $\bar{\mathcal{Y}}_m$ can be reorganized such that the resulting covariance matrix is block diagonal. This implies that blocks of M_f elements of $\bar{\mathcal{Y}}_m$ are uncorrelated. Furthermore, there is no Kronecker product among the eigenvalues of $\mathbf{C}_w$ and $\mathbf{C}_h$, which simplifies numerical implementation. Hence, we will estimate the angular parameters using $\bar{\mathcal{Y}}_m$ instead of $\mathcal{Y}_m$. Let us define the $M_o \times M_o$ diagonal matrix $\bar{\mathbf{\Lambda}}$ as

$$\bar{\mathbf{\Lambda}} = \left(\mathbf{\Lambda}_w \otimes \mathbf{I}_{M_r M_t} + \sigma_n^2 \mathbf{I}_{M_f} \otimes \mathbf{\Lambda}_h^{-1}\right) \tag{37}$$

and the $M_o \times 1$ vector $\bar{\boldsymbol{\lambda}} = \operatorname{diag}(\bar{\mathbf{\Lambda}})$. Now we can rewrite the likelihood function in (33) as

$$\begin{aligned} L(\bar{\mathcal{Y}}_1, \ldots, \bar{\mathcal{Y}}_{M_s}) \propto & -\sum_{j=1}^{M_o} \log \bar{\boldsymbol{\lambda}}_j \\ & -\frac{1}{M_s} \sum_{m=1}^{M_s} \bar{\mathcal{Y}}_m^H (\mathbf{V}_w \otimes \mathbf{I}_{M_r M_t}) \bar{\mathbf{\Lambda}}^{-1} \\ & \times \left(\mathbf{V}_w^H \otimes \mathbf{I}_{M_r M_t}\right) \bar{\mathcal{Y}}_m. \end{aligned} \tag{38}$$

Let us define the $M_o \times 1$ vector $\bar{\mathcal{X}}_m$ as

$$\bar{\mathcal{X}}_m = \left(\mathbf{V}_w^H \otimes \mathbf{I}_{M_r M_t}\right) \bar{\mathcal{Y}}_m. \tag{39}$$

An efficient form for computing $\mathcal{X}_m$ is given by

$$\bar{\mathcal{X}}_m = \operatorname{vec}\left[\left(\mathbf{V}_w^H (\operatorname{mat}(\bar{\mathcal{Y}}_m, M_r M_t, M_f))^{\mathrm{T}}\right)^{\mathrm{T}}\right]. \tag{40}$$

Finally, we can write $L(\bar{\mathcal{Y}}_1, \ldots, \bar{\mathcal{Y}}_{M_s})$ in a computationally efficient form as

$$L\left(\bar{\mathcal{Y}}_1, \ldots, \bar{\mathcal{Y}}_{M_s}\right) \propto -\sum_{j=1}^{M_o} \log \bar{\boldsymbol{\lambda}}_j - \frac{1}{M_s} \sum_{m=1}^{M_s} \mathcal{X}_m^H \bar{\mathbf{\Lambda}}^{-1} \mathcal{X}_m \tag{41}$$

where it should be noted that the multiplication between a diagonal matrix and a vector is elementwise multiplication of the main diagonal and the vector. The ML estimates of $\boldsymbol{\theta}_{wn} = \{\boldsymbol{\theta}_w, \sigma_n^2\} = \{\alpha_1, \beta_d, \tau_d, \sigma_n^2\}$ maximize the likelihood function in (41).

Even further reduction of complexity is possible in the calculation of the eigenvalues and eigenvectors of $\mathbf{C}_w$ since it is a large Toeplitz matrix. Consequently, it can be approximated by a circulant matrix [19], [20]. A circulant matrix can be diagonalized as

$$\mathbf{R}_w = \mathbf{F} \mathbf{D}_w \mathbf{F}^H \tag{42}$$

where $\mathbf{F}$ is the unitary discrete Fourier transform (DFT) matrix and $\mathbf{D}_w$ is a diagonal matrix with the eigenvalues of $\mathbf{R}_w$. Hence, $\mathbf{V}_w = \mathbf{F}$, and $\mathbf{\Lambda}_w = \mathbf{F}^H \mathbf{R}_w \mathbf{F}$. In case M_o is a power of 2, the computational complexity of $\mathbf{\Lambda}_w$ and $\bar{\mathcal{Y}}_m$ in (35) can be reduced even further by using fast Fourier transform (FFT). For other values of M_o, other algorithms that optimize the computation of the DFT can be used, like the Goertzel algorithm. There is no need to compute the off-diagonal elements of $\mathbf{F}^H \mathbf{R}_w \mathbf{F}$, and we use the following computationally efficient mapping from $\mathbf{v}(\boldsymbol{\theta}_w)$ to $\boldsymbol{\lambda}_w$ [3]:

$$\boldsymbol{\lambda}_w = T(\mathbf{v}(\boldsymbol{\theta}_w)) = \frac{1}{\sqrt{M}} \mathbf{F}^H (\mathbf{W}_1 \mathbf{v}(\boldsymbol{\theta}_w) + \mathbf{W}_2 \mathbf{v}^*(\boldsymbol{\theta}_w)) \tag{43}$$

where

$$\mathbf{W}_1 = \operatorname{diag}([M_f \quad M_f - 1 \quad \cdots \quad 1]) \tag{44}$$

and

$$\mathbf{W}_2 = \begin{bmatrix} 0 & & \cdots & & 0 \\ & & & \iddots & 1 \\ \vdots & & 0 & 2 & 0 \\ & \iddots & \iddots & \iddots & \vdots \\ 0 & M_f - 1 & 0 & \cdots & 0 \end{bmatrix}. \tag{45}$$

C. Angular-Domain Parameters

In Section II-B, we modeled the covariance matrix in angular domain, $\mathbf{C}_h$, as a function of a mixture of von Mises distributions. Assuming the number of mixture components in angular domain P is known or reliably estimated, the angular parameters are the parameters of the mixture of von Mises distributions: $\mathbf{\Theta}_h = \{\mu_1, \kappa_1, \epsilon_1, \ldots, \mu_P, \kappa_P, \epsilon_P\}$, $p = 1, \ldots, P$, with $\sum_{p=1}^{P} \epsilon_P = 1$. Due to the model in (10), the path loss is already estimated as part of the delay-domain parameters.

For the estimation of the angular domain parameters, we assume that the frequency-domain parameters are fixed. The computational complexity can be simplified as in Section III-B. Assuming $\mathbf{C}_w$ is nonsingular, let us define the transformed signal

$$\tilde{\mathcal{Y}}_m = \left(\mathbf{\Lambda}_w^{-1/2} \mathbf{F}^H \otimes \mathbf{I}_{M_r M_t}\right) \mathcal{Y}_m. \tag{46}$$

An efficient form for computation of $\tilde{\mathcal{Y}}_m$ is given by

$$\tilde{\mathcal{Y}}_m = \operatorname{vec}\left[\left(\mathbf{\Lambda}_w^{-1/2} \mathbf{F}^H (\operatorname{mat}(\tilde{\mathcal{Y}}_m, M_r M_t, M_f))^{\mathrm{T}}\right)^{\mathrm{T}}\right]. \tag{47}$$

The covariance matrix of $\tilde{\mathcal{Y}}_m$ is given by

$$\begin{aligned} \mathbf{C}_{\tilde{y}} &= \left(\mathbf{\Lambda}_w^{-1/2} \mathbf{F}^H \otimes \mathbf{I}_{M_r M_t}\right) E\left[\mathcal{Y}_m \mathcal{Y}_m^H\right] \left(\mathbf{F} \mathbf{\Lambda}_w^{-1/2} \otimes \mathbf{I}_{M_r M_t}\right) \\ &= \left(\mathbf{I}_{M_f} \otimes \mathbf{V}_h\right) \left(\mathbf{I}_{M_f} \otimes \mathbf{\Lambda}_h + \sigma_n^2 \mathbf{\Lambda}_w^{-1} \otimes \mathbf{I}_{M_r M_t}\right) \\ &\quad \times \left(\mathbf{I}_{M_f} \otimes \mathbf{V}_h^H\right). \end{aligned} \tag{48}$$

The assumption that $\mathbf{C}_w$ is nonsingular implies that $\mathbf{C}_{\tilde{y}}$ is positive definite. The covariance matrix $\mathbf{C}_{\tilde{y}}$ is block diagonal, implying that blocks of M_r elements of $\tilde{\mathcal{Y}}_m$ are uncorrelated. Also, there is no Kronecker product between the eigenvalues of $\mathbf{C}_w$ and $\mathbf{C}_h$. Consequently, we will estimate the angular parameters using $\tilde{\mathcal{Y}}_m$ instead of $\mathcal{Y}_m$. The likelihood function for $\tilde{\mathcal{Y}}_m$ is given by

$$L(\tilde{\mathcal{Y}}_1, \ldots, \tilde{\mathcal{Y}}_{M_s}) \propto -\log |\mathbf{C}_{\tilde{y}}| - \frac{1}{M_s} \sum_{m=1}^{M_s} \tilde{\mathcal{Y}}_m^H \mathbf{C}_{\tilde{y}}^{-1} \tilde{\mathcal{Y}}_m. \tag{49}$$

Let us define the $M_o \times M_o$ diagonal matrix $\tilde{\mathbf{\Lambda}}$ as

$$\tilde{\mathbf{\Lambda}} = \left(\mathbf{I}_{M_f} \otimes \mathbf{\Lambda}_h + \sigma_n^2 \mathbf{\Lambda}_w^{-1} \otimes \mathbf{I}_{M_r M_t}\right) \tag{50}$$

and the $M_o \times 1$ vector $\tilde{\boldsymbol{\lambda}} = \operatorname{diag}(\tilde{\mathbf{\Lambda}})$. Now we can write the likelihood function in (49) as

$$\begin{aligned} L(\tilde{\mathcal{Y}}_1, \ldots, \tilde{\mathcal{Y}}_{M_s}) \propto & -\sum_{j=1}^{M_o} \log \tilde{\boldsymbol{\lambda}}_j \\ & - \frac{1}{M_s} \sum_{m=1}^{M_s} \tilde{\mathcal{Y}}_m^H (\mathbf{I}_{M_f} \otimes \mathbf{V}_h) \tilde{\mathbf{\Lambda}}^{-1} \\ & \cdot \left(\mathbf{I}_{M_f} \otimes \mathbf{V}_h^H\right) \tilde{\mathcal{Y}}_m. \end{aligned} \tag{51}$$

Furthermore, let us define the $M_o \times 1$ vector $\tilde{\mathcal{X}}_m$ as

$$\tilde{\mathcal{X}}_m = \left(\mathbf{I}_{M_f} \otimes \mathbf{V}_h^H\right) \tilde{\mathcal{Y}}_m. \quad (52)$$

An efficient form for computation of $\tilde{\mathcal{X}}_m$ is given by

$$\tilde{\mathcal{X}}_m = \mathrm{vec}\left(\mathbf{V}_h^H \mathrm{mat}(\tilde{\mathcal{Y}}_m, M_r M_t, M_f)\right). \quad (53)$$

Finally, we can write $L(\tilde{\mathcal{Y}}_1, \ldots, \tilde{\mathcal{Y}}_{M_s})$ in a computationally efficient form as

$$L(\tilde{\mathcal{Y}}_1, \ldots, \tilde{\mathcal{Y}}_{M_s}) \propto -\sum_{j=1}^{M_o} \log \tilde{\boldsymbol{\lambda}}_j - \frac{1}{M_s} \sum_{m=1}^{M_s} \tilde{\mathcal{X}}_m^H \tilde{\boldsymbol{\Lambda}}^{-1} \tilde{\mathcal{X}}_m \quad (54)$$

where it should be noted again that the multiplication between a diagonal matrix and a vector is elementwise multiplication of the main diagonal and the vector. The estimates of $\theta_{h,p} = \{\kappa_p, \mu_p, \epsilon_p\}, p = 1, \ldots, P$, are those values that maximize the cost function in (54).

The optimization of the nonlinear functions in (41) and (54) can be performed using the Levenberg–Marquardt algorithm [21]. It requires the computation of the gradient and an approximation of the Hessian, which can be found in Appendix I. Also, the derivatives presented in Appendix I are used for the derivation of the performance bounds in Section IV. Efficient implementations of the gradient and the approximate Hessian can be found in Appendix II.

D. Initialization

In this section, we describe one approach for the initialization of the algorithm. The specular paths are initialized assuming $\mathbf{C}_y = \mathbf{I}$. Well-known estimators available in the literature can be used to get initial estimates for the parameters of the specular paths [2], [16].

For the diffuse scattering, the frequency-domain parameters are initialized as in [3], and assuming $\mathbf{C}_h = \mathbf{I}$.

The angular-domain parameters can be initialized using the following procedure.

1) Get initial estimate for $\mu_p, p = 0, \ldots, P-1$, using, e.g., ESPRIT assuming P sources.
2) Choose initial values for $\kappa_p, p = 0, \ldots, P-1$, uniformly drawn from the interval [0, 50], and $\epsilon_p, p = 0, \ldots, P-1$, randomly in the range [0, 1].
3) Refine the initial estimates using the method in [13], [18], with $\mathcal{Y}$ normalized by P_y, that represents the total power in $\mathsf{w}(\tau)$

$$P_y = \frac{\alpha_1}{\beta_d} e^{\beta_d \tau_d}.$$

If an arbitrary 2-D or 3-D antenna element arrangement is used, beamforming or ES-root MUSIC [22] can be applied in Step 1).

E. Summary of the Algorithm

Table I summarizes the reduced complexity algorithm for estimation of frequency- and angular-domain parameters.

We evaluate the computational complexity of the algorithm in terms of real multiplications. We assume the multiplication of two complex numbers corresponds to four real multiplications, and the multiplication of a complex and a real number corresponds to two real multiplications. Fig. 3 shows the computational complexity as a function of the number of frequency samples, M_f, assuming 20 evaluations of the cost function during the optimization, and that three cycles of the estimation procedure are used. We assume that only the diffuse components are estimated. The complexity of calculating the eigendecomposition of $\mathbf{C}_h$ is of order $\mathcal{O}((M_t M_r)^3)$ [23]. The exact number of multiplications depends on the matrix structure, but the order $\mathcal{O}((M_t M_r)^3)$ remains. Conditioning of the matrix plays a role as well.

Fig. 3 illustrates the complexities of different solutions. It is clear that the optimized method reduces the complexity by several orders of magnitude compared to the maximum-likelihood estimation, especially if FFTs are used when multiplying by $\mathbf{V}_w$.

TABLE I
REDUCED COMPLEXITY ALGORITHM FOR ESTIMATION OF FREQUENCY- AND ANGULAR-DOMAIN PARAMETERS

1) Estimate the specular component.
2) Compute $\mathbf{V}_h$ and $\boldsymbol{\Lambda}_h$.
3) Compute $\overline{\mathcal{Y}}_m$ using (35)
4) Optimize for frequency-domain parameters, (θ_w), using (41). For each iteration compute:
   a) $\boldsymbol{\Lambda}_w$ using (43)
   b) $\overline{\mathcal{X}}_m$ using (40) and FFT.
5) Compute $\overline{\mathcal{Y}}_m$ using (47) and FFT.
6) Optimize for angular-domain parameters, (Θ_h), using (54). For each iteration compute:
   a) $\mathbf{V}_h$ and $\boldsymbol{\Lambda}_h$
   b) $\tilde{\mathcal{X}}_m$ using (53).
7) Repeat 1)–6) until convergence or a maximum number of iterations is achieved.

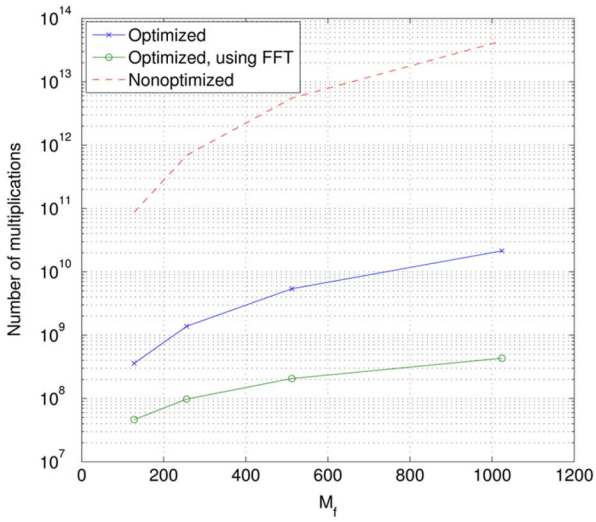

Fig. 3. Comparison of computational complexity of direct optimization of likelihood function and the optimized algorithm as a function of the number of frequency samples, M_f. The optimized method reduces the complexity by several orders of magnitude compared to the maximum-likelihood estimation, especially if FFT is used when multiplying by $\mathbf{V}_w$.

The effectiveness of this reduction depends on the number of cycles. A reasonable criterion for convergence is to stop the algorithm if after any two iterations the relative change for all parameters is less than a predefined threshold, e.g., 10^{-2}. For this threshold value, the parameters usually converge in less than five iterations for all versions of the algorithm.

IV. Performance Bounds

The $\{i,k\}$th element of the Fisher information matrix (FIM) for a circular complex white Gaussian variable are given by [15]

$$\mathcal{I}_{ik} = M_s 2\Re\left\{\left(\frac{\partial \mathbf{s}^H}{\partial\theta_i}\right)\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{s}}{\partial\theta_k}\right)\right\} + M_s \mathrm{tr}\left\{\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{C}_y}{\partial\theta_i}\right)\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{C}_y}{\partial\theta_k}\right)\right\}. \quad (55)$$

Hence, the FIM is block diagonal if $\boldsymbol{\theta}_{\mathrm{sp}}$ and $\boldsymbol{\theta}_{\mathrm{dn}}$ contain uncorrelated parameters, i.e.,

$$\mathcal{I} = \begin{bmatrix} \mathcal{I}(\boldsymbol{\theta}_{\mathrm{sp}}) & \mathbf{0} \\ \mathbf{0} & \mathcal{I}(\boldsymbol{\theta}_{\mathrm{dn}}) \end{bmatrix} \quad (56)$$

where $\mathcal{I}(\boldsymbol{\theta}_{\mathrm{sp}})$ is the FIM for the specular components, $\boldsymbol{\theta}_{\mathrm{dn}} = \{\boldsymbol{\theta}_{\mathrm{wn}}, \boldsymbol{\Theta}_{\mathrm{h}}\}$, and $\mathcal{I}(\boldsymbol{\theta}_{\mathrm{dn}})$ is the FIM for the diffuse scattering parameters plus noise variance. As a consequence, its inverse is also block diagonal [23]. This means that the parameters of the specular and diffuse components are asymptotically decoupled, and the respective CRLBs can be derived independently. The CRLB for the parameters of the specular components can be found in [2], [3]. The CRLB for the parameters of the diffuse scattering components is given by (55) as

$$\mathcal{I}_{ik}(\boldsymbol{\theta}_{\mathrm{dn}}) = M_s \mathrm{tr}\left\{\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{C}_y}{\partial\theta_i}\right)\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{C}_y}{\partial\theta_k}\right)\right\}. \quad (57)$$

Let us define the matrices containing the partial derivatives with respect to all N_p elements of the parameter vector $\boldsymbol{\theta}_{\mathrm{dn}}$, as (58) and (59), shown at the bottom of the page. Using (58) and (59), the FIM can be expressed in a compact form as

$$\mathcal{I} = M_s \mathbf{D}_1^{\mathrm{T}} \mathbf{D}_2. \quad (60)$$

The partial derivatives of $\mathbf{C}_y$ with respect to the diffuse scattering component parameters is given by

$$\frac{\partial \mathbf{C}_y}{\partial\theta_i} = \frac{\partial \mathbf{C}_w}{\partial\theta_i} \otimes \mathbf{C}_h, \qquad \theta_i \in \boldsymbol{\theta}_w \quad (61)$$

$$\frac{\partial \mathbf{C}_y}{\partial\theta_i} = \mathbf{C}_w \otimes \frac{\partial \mathbf{C}_h}{\partial\theta_i}, \qquad \theta_i \in \boldsymbol{\Theta}_h \quad (62)$$

$$\frac{\partial \mathbf{C}_y}{\partial\sigma^2} = \mathbf{I}. \quad (63)$$

The partial derivatives of $\mathbf{C}_w$ and $\mathbf{C}_h$ with respect to the propagation parameters can be found in Appendix I.

V. Simulation Results

In this section, simulation examples are presented in order to illustrate the performance of the described parameter estimation procedure. The receiver is equipped with a ULA having $M_r = 8$ antennas and the transmitter uses $M_t = 1$ antenna. The number of frequency points is $M_f = 128$, and the number of channel realizations is $M_s = 5$. For the frequency-domain parameters, typical values often observed in channel sounding experiments are used: $\sigma_n^2 = 0.1$, $\alpha_1 = 1$, $\beta = 0.07$, and $\tau = 0.1$. The angular-domain parameters are defined as $\phi = \{60^\circ, 120^\circ\}$, $\kappa = \{10, 50\}$, $\epsilon = \{0.4, 0.6\}$, corresponding to two clusters in the angular domain. One specular component is assumed to be present, and it is modeled as

$$\mathbf{s}(k) = \gamma \mathbf{b}(\varphi_{\mathrm{R}}) \exp(-j2\pi k\tau) \quad (64)$$

where γ is the complex gain, $\mathbf{b}(\varphi_{\mathrm{R}})$ is the steering vector for receive azimuth angle φ_{R}, and τ is the normalized delay. For the simulation, the values are set as $\gamma = 0.8e^{j*\pi/5}$, $\varphi_{\mathrm{R}} = 80^\circ$, and $\tau = 0.12$.

The received signal is generated as

$$\mathbf{y}(k) = \mathbf{s}(k) + \mathbf{R}^{1/2}\mathbf{n}_2(k) + \mathbf{n}(k) \quad (65)$$

where $\mathbf{n}_2(k)$ is a circular complex white Gaussian process and $\mathbf{R}^{1/2}$ is obtained by the Cholesky decomposition of $\mathbf{C}_w \otimes \mathbf{C}_h$. This implies that the covariance matrix of $\mathbf{R}^{1/2}\mathbf{n}_2(k)$ is given by $\mathbf{C}_w \otimes \mathbf{C}_h$. The vector $\mathbf{n}(k)$ is a circular complex white Gaussian process representing the measurement noise.

The iterative procedure described in Section III is repeated five times, starting with the estimation of the specular component and proceeding as illustrated in Fig. 2. The model order, i.e., number of clusters and specular paths, is assumed known. The frequency-domain parameters are computed using the approximation of a Toeplitz matrix as a circulant matrix described in Section III-B. The parameters are initialized as described in Section III-D. In Figs. 4 and 5, we compare the power delay profile (PDP) and power angular profile (PAP) obtained using the estimation procedure described in this article with the actual PDP and PAP, respectively. The curves overlap almost perfectly. In the example, $M_s \ll M_o$, i.e., the full sample covariance matrix is rank deficient. Still, the estimator is able to provide high-precision estimates for the time-delay distribution, angular distribution, and specular component. The PAP is compared to the output of the Bartlett beamformer, showing the gain in using the combined procedure to estimate both signal components iteratively. The beamformer is only able to estimate the angle of the specular component, but it does not provide any useful information about the diffuse scattering component.

$$\mathbf{D}_1 = \left[\mathrm{vec}\left\{\left(\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{C}_y}{\partial(\boldsymbol{\theta}_{\mathrm{dn}})_1}\right)\right)^{\mathrm{T}}\right\} \quad \cdots \quad \mathrm{vec}\left\{\left(\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{C}_y}{\partial(\boldsymbol{\theta}_{\mathrm{dn}})_{N_p}}\right)\right)^{\mathrm{T}}\right\}\right] \quad (58)$$

$$\mathbf{D}_2 = \left[\mathrm{vec}\left\{\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{C}_y}{\partial(\boldsymbol{\theta}_{\mathrm{dn}})_1}\right)\right\} \quad \cdots \quad \mathrm{vec}\left\{\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{C}_y}{\partial(\boldsymbol{\theta}_{\mathrm{dn}})_{N_p}}\right)\right\}\right]. \quad (59)$$

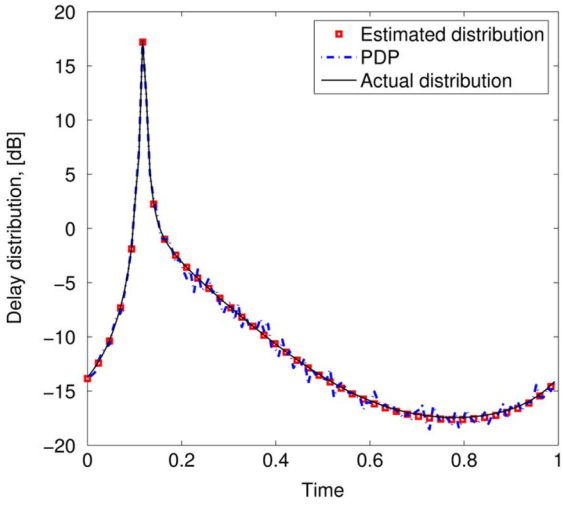

Fig. 4. Comparison of estimated power delay profile and actual power delay profile. The curves overlap almost perfectly. The specular component is identified as a sharp peak at $\tau_n = 0.12$, while the diffuse component corresponds to the exponential curve.

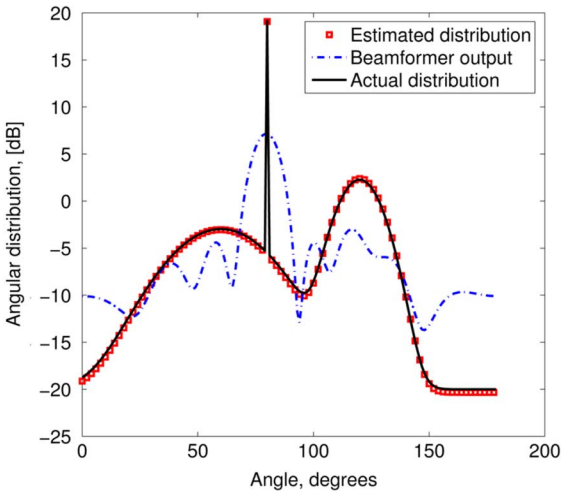

Fig. 5. Comparison of estimated power angular profile and actual power angular profile. The curves overlap almost perfectly. Also shown is the output of the Bartlett beamformer. The specular component is identified as a sharp peak at $\varphi = 80°$, while the diffuse component corresponds to the mixture of von Mises distributions.

The algorithms that approximate the mean angle and angular variance using two separate paths around the mean are called Spread F [10], where F denotes the underlying algorithm to estimate the paths. Table II shows the estimates obtained using the proposed method and Spread ESPRIT [10]. The results are an average over 300 runs. The parameters of the DSC and specular components are the same as in the previous simulation. The angular spread is shown in degrees using the mapping $\sigma_\mu \approx \kappa^{-1/2}$ [14]. Hence, for the simulated values of $\kappa = \{10, 50\}$, we obtain $\sigma_\mu = \{18.12°, 8.10°\}$.

The estimator proposed in this paper overperforms Spread ESPRIT for all parameters of the diffuse component, while presenting similar performance for the specular component. One problem for the application of the Spread F techniques is that it is very difficult in a real-world environment to identify which

TABLE II
COMPARISON BETWEEN THE PROPOSED METHOD AND SPREAD ESPRIT

	Correct values	Proposed method	Spread ESPRIT
μ_1	$60°$	$59.98°$	$56.66°$
$\sigma_{\mu,1}$	$18.12°$	$18.14°$	$9.20°$
μ_2	$120°$	$119.99°$	$119.15°$
$\sigma_{\mu,2}$	$8.10°$	$8.07°$	$3.09°$
φ	$80°$	$79.99°$	$79.97°$

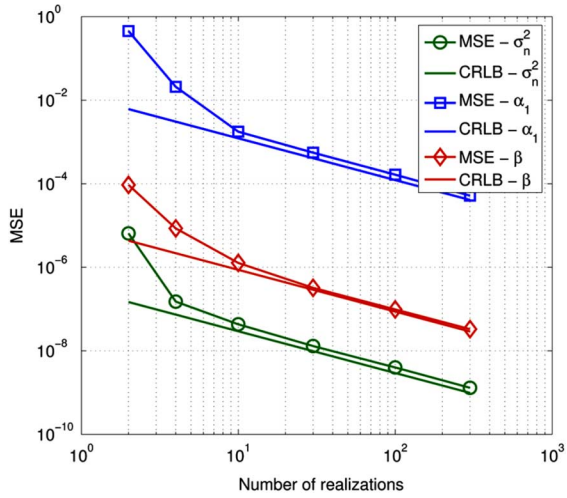

Fig. 6. CRLB of frequency-domain parameters $\{\alpha_1, \beta_n\}$ and noise variance as a function of the number of channel realizations, M_s. The MSE after two iterations of the estimation procedure is shown for comparison.

of the identified waves belong to a distributed scatterer in particular and which one is a specular component. This issue limits the application of Spread F techniques to well separated sources with small angular spread, as already noted in [10].

In Figs. 6–10, we compare the mean-square error (MSE) of the estimates after two cycles with the CRLB as a function of the number of channel realizations M_s. The angular-domain parameters are set to $\phi = \{50°, 100°\}$, $\kappa = \{5, 150\}$, $\epsilon = \{0.4, 0.6\}$, corresponding to two clusters in the angular domain. The frequency-domain parameters remain unchanged. No specular components are present. It can be observed that all parameters converge close to the CRLB for a relatively small sample size. The exception is the relative delay τ_n, which presents a noticeable gap with respect to its CRLB. However, no bias is observed, since the curve is parallel to the CRLB.

In order to verify the robustness of the algorithm, we apply the estimator to data that do not follow exactly the assumptions used for its derivation. We generate the data as a sum of individual propagation paths

$$\mathbf{y}(k) = \sum_{l=1}^{L} \mathbf{s}_l(k) + \mathbf{n}(k) \tag{66}$$

where L is the number of individual paths, and $\mathbf{s}_l(k)$ are defined as in (2). For this simulation, we use $L = 24$, the noise variance is 0.01, and $M_s = 4$. Convergence is achieved after four iterations, assuming the parameters converge after changing by less than 5×10^{-3}. The specular paths are not estimated individually, but rather the joint angle-delay distribution of diffuse scattering

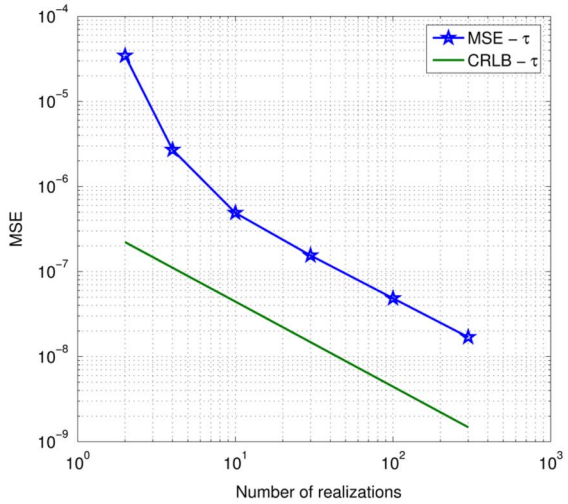

Fig. 7. CRLB of base delay as a function of the number of channel realizations M_s. The MSE after two iterations of the estimation procedure is shown for comparison.

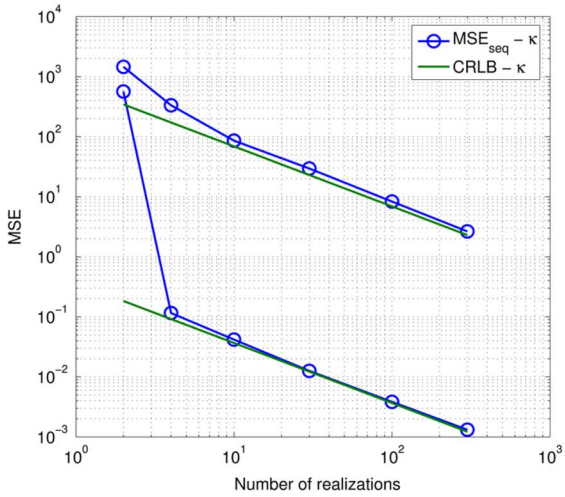

Fig. 9. CRLB of dispersion parameter as a function of the number of channel realizations M_s. The MSE after two iterations of the estimation procedure is shown for comparison.

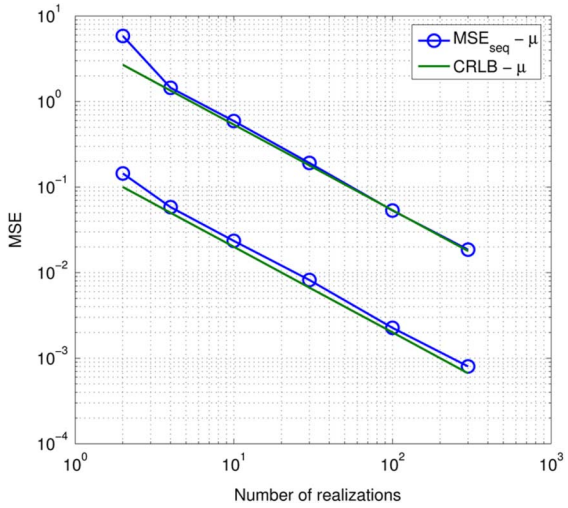

Fig. 8. CRLB of mean angle as a function of the number of channel realizations M_s. The MSE after two iterations of the estimation procedure is shown for comparison.

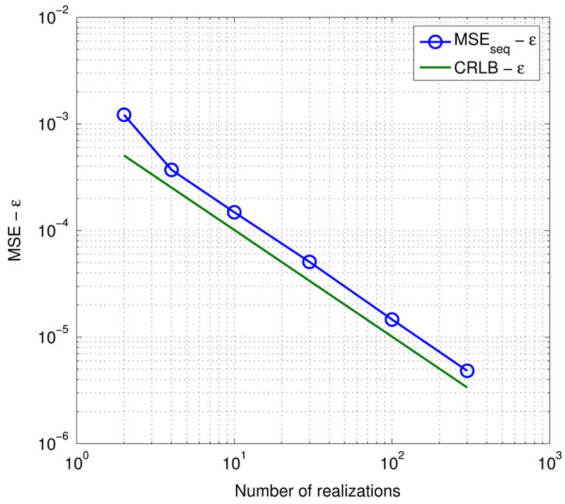

Fig. 10. CRLB of mixture proportion of the first angular cluster as a function of the number of channel realizations, M_s. The MSE after two iterations of the estimation procedure is shown for comparison. The maximum number of mixture components is assumed to be known.

is used to characterize the data, assuming $P = 1$. Fig. 11 shows the estimated joint angle-delay distribution and the individual paths used for generating the data. It can be observed that the estimated distributions provide a good fit for the data.

VI. Conclusion

In this paper, we derived a joint estimator for the parameters of the concentrated propagation paths and the distributed scattering component that are frequently found in MIMO radio propagation measurements. In particular, we estimate the parameters of the scattering component in both spatial and temporal domains. The power-delay profile of the scattering component is modeled using an exponential distribution, which is typically observed in measurement campaigns. The power angular profile is modeled using a mixture of angular von Mises distributions. The simulation results show that this procedure converges to the estimates of both specular and diffuse components with high fidelity. Convergence is achieved with few iterations.

We derived computationally efficient methods for finding the approximate ML estimates. The structure of the covariance matrices is fully exploited. Complexity studies show that the reduction in the number of real multiplications is approximately three to five orders of magnitude. We also present computationally efficient methods to compute the gradients and Hessians, which are useful for the implementation of the optimization routines.

Furthermore, we derived the CRLB for the problem and showed that the variance of the estimates converges close to the bound for a relatively small number of cycles of the estimation procedure and small number of channel realizations (small sample size). For some parameters the CRLB is not attained, but no error floor indicating bias is observed.

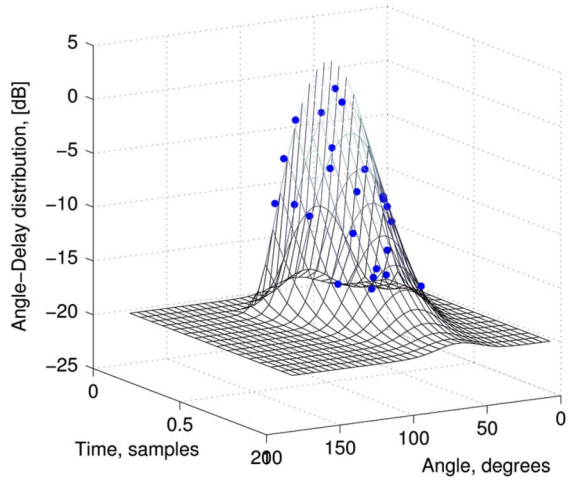

Fig. 11. Estimated joint angle-delay distribution and the actual individual paths used to generate the data, denoted by the circles. The estimated distributions provide a good fit for the data.

Appendix I
Gradient and Approximate Hessian

The derivative of (38) with respect to the ith element of $\boldsymbol{\theta}_w = \{\alpha_1, \beta_d, \tau_d\}$ is given by

$$\begin{aligned}\frac{\partial L(\bar{\mathcal{Y}})}{\partial(\boldsymbol{\theta}_w)_i} &= -\mathrm{tr}\left(\mathbf{C}_{\bar{y}}^{-1}\frac{\partial \mathbf{C}_{\bar{y}}}{\partial(\boldsymbol{\theta}_w)_i}\right) \\ &\quad + \frac{1}{M_s}\sum_{m=1}^{M_s}\bar{\mathcal{Y}}_m^H\mathbf{C}_{\bar{y}}^{-1}\frac{\partial \mathbf{C}_{\bar{y}}}{\partial(\boldsymbol{\theta}_w)_i}\mathbf{C}_{\bar{y}}^{-1}\bar{\mathcal{Y}}_m \\ &= -\mathrm{tr}\left(\mathbf{C}_{\bar{y}}^{-1}\left(\frac{\partial \mathbf{C}_w}{\partial(\boldsymbol{\theta}_w)_i}\otimes\mathbf{I}\right)\right) \\ &\quad + \frac{1}{M_s}\sum_{m=1}^{M_s}\bar{\mathcal{Y}}_m^H\mathbf{C}_{\bar{y}}^{-1}\left(\frac{\partial \mathbf{C}_w}{\partial(\boldsymbol{\theta}_w)_i}\otimes\mathbf{I}\right)\mathbf{C}_{\bar{y}}^{-1}\bar{\mathcal{Y}}_m. \quad (67)\end{aligned}$$

The derivative of $\mathbf{C}_w$ with respect to the frequency-domain parameters is given by

$$\frac{\partial \mathbf{C}_w}{\partial(\boldsymbol{\theta}_w)_i} = \mathrm{toep}\left(\frac{\partial \mathbf{v}(\boldsymbol{\theta}_w)}{\partial(\boldsymbol{\theta}_w)_i}, \frac{\partial \mathbf{v}(\boldsymbol{\theta}_w)}{\partial(\boldsymbol{\theta}_w)_i}^H\right) \quad (68)$$

and $(\partial \mathbf{v}(\boldsymbol{\theta}_w))/(\partial(\boldsymbol{\theta}_w)_i)$ is calculated from (13) as (69)–(71), shown at the bottom of the page. From (36), the derivative of $\mathbf{C}_{\bar{y}}$ with respect to the noise variance σ_n^2 is given by

$$\frac{\partial \mathbf{C}_{\bar{y}}}{\partial \sigma_n^2} = \mathbf{I}\otimes\boldsymbol{\Lambda}_h^{-1} \quad (72)$$

and hence (67) simplifies to

$$\begin{aligned}\frac{\partial L(\bar{\mathcal{Y}})}{\partial \sigma_n^2} &= -\mathrm{tr}\left(\mathbf{C}_{\bar{y}}^{-1}\frac{\partial \mathbf{C}_{\bar{y}}}{\partial \sigma_n^2}\right) \\ &\quad + \frac{1}{M_s}\sum_{m=1}^{M_s}\bar{\mathcal{Y}}_m^H\mathbf{C}_{\bar{y}}^{-1}\frac{\partial \mathbf{C}_{\bar{y}}}{\partial \sigma_n^2}\mathbf{C}_{\bar{y}}^{-1}\bar{\mathcal{Y}}_m \\ &= -\mathrm{tr}\left(\mathbf{C}_{\bar{y}}^{-1}\left(\mathbf{I}\otimes\boldsymbol{\Lambda}_h^{-1}\right)\right) \\ &\quad + \frac{1}{M_s}\sum_{m=1}^{M_s}\bar{\mathcal{Y}}_m^H\mathbf{C}_{\bar{y}}^{-1}\left(\mathbf{I}\otimes\boldsymbol{\Lambda}_h^{-1}\right)\mathbf{C}_{\bar{y}}^{-1}\bar{\mathcal{Y}}_m. \quad (73)\end{aligned}$$

An approximation of the Hessian is given by the FIM, which gives the expected value of the Hessian. Using the fact that $\bar{\mathcal{Y}}$ is a circular complex white Gaussian variable, we obtain the partial derivative with respect to the ith and kth elements of $\boldsymbol{\theta}_{wn} = \{\alpha_1, \beta_d, \tau_d, \sigma_n^2\}$ as

$$\begin{aligned}&E\left[\frac{\partial^2 L(\bar{\mathcal{Y}})}{\partial(\boldsymbol{\theta}_{wn})_i\partial(\boldsymbol{\theta}_{wn})_k}\right] \\ &\quad = M_s\mathrm{tr}\left\{\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{C}_y}{\partial(\boldsymbol{\theta}_{wn})_i}\right)\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{C}_y}{\partial(\boldsymbol{\theta}_{wn})_k}\right)\right\}. \quad (74)\end{aligned}$$

Recalling that the trace of the product between two square matrices can be expressed as

$$\mathrm{tr}(\mathbf{AB}) = \mathrm{vec}(\mathbf{A}^{\mathrm{T}})^{\mathrm{T}}\cdot\mathrm{vec}(\mathbf{B}) \quad (75)$$

we can write

$$\begin{aligned}&E\left[\frac{\partial^2 L(\bar{\mathcal{Y}})}{\partial(\boldsymbol{\theta}_{wn})_i\partial(\boldsymbol{\theta}_{wn})_k}\right] \\ &\quad = M_s\mathrm{vec}\left\{\left(\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{C}_y}{\partial(\boldsymbol{\theta}_{wn})_i}\right)\right)^{\mathrm{T}}\right\}^{\mathrm{T}} \\ &\qquad \times \mathrm{vec}\left\{\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{C}_y}{\partial(\boldsymbol{\theta}_{wn})_k}\right)\right\}. \quad (76)\end{aligned}$$

Let us define the matrices $\mathbf{D}_{1,wn}$ and $\mathbf{D}_{2,wn}$ containing the partial derivatives with respect to all N_w elements of the parameter

$$\frac{\partial \mathbf{v}(\boldsymbol{\theta}_w)}{\partial \alpha_1} = \frac{1}{M_f}\left[\frac{1}{\beta_d} \quad \cdots \quad \frac{e^{-j2\pi(M_f-1)\tau_d}}{\beta_d + j2\pi\frac{M_f-1}{M_f}}\right] \quad (69)$$

$$\frac{\partial \mathbf{v}(\boldsymbol{\theta}_w)}{\partial \beta_d} = -\frac{\alpha_1}{M_f}\left[\frac{1}{(\beta_d)^2} \quad \cdots \quad \frac{e^{-j2\pi(M_f-1)\tau_d}}{\left(\beta_d + j2\pi\frac{M_f-1}{M_f}\right)^2}\right] \quad (70)$$

and

$$\frac{\partial \mathbf{v}(\boldsymbol{\theta}_w)}{\partial \tau_d} = \frac{\alpha_1}{M_f}\left[0 \quad \frac{-j2\pi e^{-j2\pi\tau_d}}{\beta_d + j2\pi\frac{M_f-1}{M_f}} \quad \cdots \quad \frac{-j2\pi(M_f-1)e^{-j2\pi(M_f-1)\tau_d}}{\beta_d + j2\pi\frac{M_f-1}{M_f}}\right]. \quad (71)$$

vector $\boldsymbol{\theta}_w$ and the partial derivative with respect to the noise variance σ_n^2 as (77) and (78), shown at the bottom of the page. Hence, the approximation of the Hessian can be expressed in a compact form as

$$E\left[\frac{\partial^2 L(\bar{\mathcal{Y}})}{\partial(\boldsymbol{\theta}_{wn})_i \partial(\boldsymbol{\theta}_{wn})_k}\right] = M_s \mathbf{D}_{1,wn}^{\mathrm{T}} \mathbf{D}_{2,wn}. \tag{79}$$

The derivative of (51) with respect to the angular-domain parameters is given by

$$\begin{aligned}\frac{\partial L(\tilde{\mathcal{Y}})}{\partial(\boldsymbol{\theta}_h)_i} &= -\mathrm{tr}\left(\mathbf{C}_{\tilde{y}}^{-1}\frac{\partial \mathbf{C}_{\tilde{y}}}{\partial(\boldsymbol{\theta}_h)_i}\right) \\ &\quad + \frac{1}{M_s}\sum_{m=1}^{M_s} \tilde{\mathcal{Y}}_m^H \mathbf{C}_{\tilde{y}}^{-1}\frac{\partial \mathbf{C}_{\tilde{y}}}{\partial(\boldsymbol{\theta}_h)_i}\mathbf{C}_{\tilde{y}}^{-1}\tilde{\mathcal{Y}}_m \\ &= -\mathrm{tr}\left(\mathbf{C}_{\tilde{y}}^{-1}\left(\mathbf{I}\otimes\frac{\partial \mathbf{C}_h}{\partial(\boldsymbol{\theta}_h)_i}\right)\right) \\ &\quad + \frac{1}{M_s}\sum_{m=1}^{M_s} \tilde{\mathcal{Y}}_m^H \mathbf{C}_{\tilde{y}}^{-1}\left(\mathbf{I}\otimes\frac{\partial \mathbf{C}_h}{\partial(\boldsymbol{\theta}_h)_i}\right)\mathbf{C}_{\tilde{y}}^{-1}\tilde{\mathcal{Y}}_m. \end{aligned}\tag{80}$$

The $\{m_1, m_2\}$th element of the derivative of $\mathbf{C}_h$ with respect to the angular-domain parameters for the SVA model in (20) is given by

$$\begin{aligned}&\left\{\frac{\partial \mathbf{C}_h}{\partial \mu_p}\right\}_{m_1,m_2} \\ &\quad = -\epsilon_p I_0^{-1}(\kappa_p) I_1(\beta_p)\beta_p^{-1}\kappa_p b_{m_1 m_2}\sin\mu_p, \end{aligned}\tag{81}$$

$$\begin{aligned}&\left\{\frac{\partial \mathbf{C}_h}{\partial \kappa_p}\right\}_{m_1,m_2} \\ &\quad = \epsilon_p\left[-I_0^{-2}(\kappa_p)I_1(\kappa_p)I_0(\beta_p) + \right. \\ &\qquad \left. + I_0^{-1}(\kappa_p)I_1(\beta_p)\beta_p^{-1}(\kappa_p + b_{m_1 m_2}\cos\mu_p)\right]. \end{aligned}\tag{82}$$

The condition $\sum_{p=1}^{P}\epsilon_p = 1$ can be rewritten as $\epsilon_P = 1 - \sum_{p=1}^{P-1}\epsilon_p$, and then we can calculate the derivatives of $\mathbf{C}_y(\boldsymbol{\theta})$ with respect to the mixture proportions $\epsilon_p, p = 1, \ldots, P-1$, as

$$\left\{\frac{\partial \mathbf{C}_h}{\partial \epsilon_p}\right\}_{m_1,m_2} = I_0^{-1}(\kappa_p)I_0(\beta_p) - I_0^{-1}(\kappa_P)I_0(\beta_P). \tag{83}$$

The derivative of $\mathbf{C}_h$ with respect to the angular-domain parameters for the one-ring model can be found in [18].

Similarly to the frequency-domain and noise parameters, an approximation of the Hessian for estimation of the angular-domain parameters is given by

$$\begin{aligned}&E\left[\frac{\partial^2 L(\tilde{\mathcal{Y}})}{\partial(\boldsymbol{\theta}_h)_i \partial(\boldsymbol{\theta}_h)_k}\right] \\ &\quad = M_s \mathrm{tr}\left\{\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{C}_y}{\partial(\boldsymbol{\theta}_h)_i}\right)\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{C}_y}{\partial(\boldsymbol{\theta}_h)_k}\right)\right\} \\ &\quad = M_s \mathrm{vec}\left\{\left(\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{C}_y}{\partial(\boldsymbol{\theta}_h)_i}\right)\right)^{\mathrm{T}}\right\}^{\mathrm{T}} \\ &\qquad \times \mathrm{vec}\left\{\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{C}_y}{\partial(\boldsymbol{\theta}_h)_k}\right)\right\}. \end{aligned}\tag{84}$$

Let us define the matrices containing the partial derivatives with respect to all N_h elements of the parameter vector $\boldsymbol{\theta}_h$ as (85) and (86), shown at the bottom of the page. Hence, the approximation of the Hessian can be expressed in a compact form as

$$E\left[\frac{\partial^2 L(\tilde{\mathcal{Y}})}{\partial(\boldsymbol{\theta}_h)_i \partial(\boldsymbol{\theta}_h)_k}\right] = M_s \mathbf{D}_{1,h}^{\mathrm{T}} \mathbf{D}_{2,h}. \tag{87}$$

Efficient formulas for computing (67), (73), (77), (78), (80), (85), and (86) can be found in Appendix II.

Appendix II
Efficient Implementations of Matrix Operations

In order to derive an efficient implementation for (67), we define

$$\begin{aligned}A_1 &= -\mathrm{tr}\left(\mathbf{C}_{\bar{y}}^{-1}\left(\frac{\partial \mathbf{C}_w}{\partial(\boldsymbol{\theta}_w)_i}\otimes\mathbf{I}\right)\right) \\ &= -\mathrm{tr}\left((\mathbf{V}_w\otimes\mathbf{I}_{M_r M_t})\bar{\boldsymbol{\Lambda}}^{-1}\left(\mathbf{V}_w^H\otimes\mathbf{I}_{M_r M_t}\right)\right. \\ &\qquad \left.\times\left(\frac{\partial \mathbf{C}_w}{\partial(\boldsymbol{\theta}_w)_i}\otimes\mathbf{I}\right)\right) \\ &= -\mathrm{tr}\left(\bar{\boldsymbol{\Lambda}}^{-1}\left(\mathbf{V}_w^H\frac{\partial \mathbf{C}_w}{\partial(\boldsymbol{\theta}_w)_i}\mathbf{V}_w\otimes\mathbf{I}\right)\right) \end{aligned}\tag{88}$$

$$\mathbf{D}_{1,wn} = \left[\mathrm{vec}\left\{\left(\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{C}_w}{\partial(\boldsymbol{\theta}_{wn})_1}\otimes\mathbf{I}\right)\right)^{\mathrm{T}}\right\} \quad \cdots \quad \mathrm{vec}\left\{\left(\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{C}_w}{\partial(\boldsymbol{\theta}_{wn})_{N_w}}\otimes\mathbf{I}\right)\right)^{\mathrm{T}}\right\} \quad \mathrm{vec}\left\{\left(\mathbf{C}_y^{-1}\left(\mathbf{I}\otimes\boldsymbol{\Lambda}_h^{-1}\right)\right)^{\mathrm{T}}\right\}\right] \tag{77}$$

$$\mathbf{D}_{2,wn} = \left[\mathrm{vec}\left\{\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{C}_w}{\partial(\boldsymbol{\theta}_{wn})_1}\otimes\mathbf{I}\right)\right\} \quad \cdots \quad \mathrm{vec}\left\{\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{C}_w}{\partial(\boldsymbol{\theta}_{wn})_{N_w}}\otimes\mathbf{I}\right)\right\} \quad \mathrm{vec}\left\{\mathbf{C}_y^{-1}\left(\mathbf{I}\otimes\boldsymbol{\Lambda}_h^{-1}\right)\right\}\right]. \tag{78}$$

$$\mathbf{D}_{1,h} = \left[\mathrm{vec}\left\{\left(\mathbf{C}_y^{-1}\left(\mathbf{I}\otimes\frac{\partial \mathbf{C}_h}{\partial(\boldsymbol{\theta}_h)_1}\right)\right)^{\mathrm{T}}\right\} \quad \cdots \quad \mathrm{vec}\left\{\left(\mathbf{C}_y^{-1}\left(\mathbf{I}\otimes\frac{\partial \mathbf{C}_h}{\partial(\boldsymbol{\theta}_h)_{N_h}}\right)\right)^{\mathrm{T}}\right\}\right] \tag{85}$$

$$\mathbf{D}_{2,h} = \left[\mathrm{vec}\left\{\mathbf{C}_y^{-1}\left(\mathbf{I}\otimes\frac{\partial \mathbf{C}_h}{\partial(\boldsymbol{\theta}_h)_1}\right)\right\} \quad \cdots \quad \mathrm{vec}\left\{\mathbf{C}_y^{-1}\left(\mathbf{I}\otimes\frac{\partial \mathbf{C}_h}{\partial(\boldsymbol{\theta}_h)_{N_h}}\right)\right\}\right]. \tag{86}$$

where $(\boldsymbol{\theta}_w)_i$ is the ith element of $\boldsymbol{\theta}_w = \{\alpha_1, \beta_d, \tau_d\}$. It is assumed in Section III-B that the eigenvectors of a Toeplitz matrix are approximated by the DFT matrix, and $\mathbf{V}_w^H((\partial \mathbf{C}_w)/(\partial(\boldsymbol{\theta}_w)_i))\mathbf{V}_w$ is also a diagonal matrix, given by $\mathcal{T}((\partial \mathbf{v}(\boldsymbol{\theta}_w))/(\partial(\boldsymbol{\theta}_w)_i))$, with $\mathcal{T}(\cdot)$ defined in (43). Since $\bar{\boldsymbol{\Lambda}}^{-1}$ and $(\mathbf{V}_w^H(\partial \mathbf{C}_w)/(\partial(\boldsymbol{\theta}_w)_i)\mathbf{V}_w \otimes \mathbf{I})$ are diagonal matrices, (88) can be written as

$$A_1 = -\mathrm{diag}(\bar{\boldsymbol{\Lambda}}^{-1})^{\mathrm{T}}\mathrm{vec}\left\{\begin{bmatrix} \mathcal{T}^{\mathrm{T}}\left(\frac{\partial \mathbf{v}(\boldsymbol{\theta}_w)}{\partial(\boldsymbol{\theta}_w)_i}\right) \\ \vdots \\ \mathcal{T}^{\mathrm{T}}\left(\frac{\partial \mathbf{v}(\boldsymbol{\theta}_w)}{\partial(\boldsymbol{\theta}_w)_i}\right) \end{bmatrix}\right\}. \tag{89}$$

We also define

$$\begin{aligned} B_1 &= \frac{1}{M_s}\sum_{m=1}^{M_s} \bar{\mathcal{Y}}_m^H \mathbf{C}_{\bar{y}}^{-1}\left(\frac{\partial \mathbf{C}_w}{\partial(\boldsymbol{\theta}_w)_i}\otimes \mathbf{I}\right)\mathbf{C}_{\bar{y}}^{-1}\bar{\mathcal{Y}}_m \\ &= \frac{1}{M_s}\sum_{m=1}^{M_s} \bar{\mathcal{Y}}_m^H (\mathbf{V}_w \otimes \mathbf{I}_{M_rM_t})\bar{\boldsymbol{\Lambda}}^{-1} \\ &\quad \times \left(\mathbf{V}_w^H \otimes \mathbf{I}_{M_rM_t}\right)\left(\frac{\partial \mathbf{C}_w}{\partial(\boldsymbol{\theta}_w)_i}\otimes \mathbf{I}\right) \\ &\quad \times (\mathbf{V}_w \otimes \mathbf{I}_{M_rM_t})\bar{\boldsymbol{\Lambda}}^{-1}\left(\mathbf{V}_w^H \otimes \mathbf{I}_{M_rM_t}\right)\bar{\mathcal{Y}}_m \\ &= \frac{1}{M_s}\sum_{m=1}^{M_s} (\bar{\boldsymbol{\Lambda}}^{-1}\bar{\mathcal{X}}_m)^H\left(\mathbf{V}_w^H\frac{\partial \mathbf{C}_w}{\partial(\boldsymbol{\theta}_w)_i}\mathbf{V}_w \otimes \mathbf{I}\right)\bar{\boldsymbol{\Lambda}}^{-1}\bar{\mathcal{X}}_m \\ &= \frac{1}{M_s}\sum_{m=1}^{M_s} (\mathrm{diag}(\bar{\boldsymbol{\Lambda}}^{-1})\circ \bar{\mathcal{X}}_m)^H \\ &\quad \times \mathrm{vec}\left[\left(\mathrm{diag}\left(\mathcal{T}\left(\frac{\partial \mathbf{v}(\boldsymbol{\theta}_w)}{\partial(\boldsymbol{\theta}_w)_i}\right)\right)\right.\right. \\ &\qquad \left.\left. \times \mathrm{mat}(\mathrm{diag}(\bar{\boldsymbol{\Lambda}}^{-1})\circ \bar{\mathcal{X}}_m, M_rM_t, M_f)^{\mathrm{T}}\right)^{\mathrm{T}}\right] \end{aligned} \tag{90}$$

where $\circ$ denotes elementwise multiplication and was used the definition of $\bar{\mathcal{X}}_m$ in (39) as well as its corresponding efficient implementation (40). The $\mathrm{mat}(\cdot, M, N)$ reshapes a vector into an $M \times N$ matrix, as defined in (29).

Hence, (67) can be implemented efficiently as

$$\frac{\partial L(\bar{\mathcal{Y}})}{\partial(\boldsymbol{\theta}_w)_i} = A_1 + B_1. \tag{91}$$

Similarly, in order to derive an efficient implementation for (73), we define

$$\begin{aligned} A_2 &= -\mathrm{tr}\left(\mathbf{C}_{\bar{y}}^{-1}\frac{\partial \mathbf{C}_{\bar{y}}}{\partial \sigma_n^2}\right) \\ &= -\mathrm{tr}\left((\mathbf{V}_w \otimes \mathbf{I}_{M_rM_t})\bar{\boldsymbol{\Lambda}}^{-1}(\mathbf{V}_w^H \otimes \mathbf{I}_{M_rM_t})\left(\mathbf{I}\otimes \boldsymbol{\Lambda}_h^{-1}\right)\right) \\ &= -\mathrm{tr}\left(\bar{\boldsymbol{\Lambda}}^{-1}\left(\mathbf{I}\otimes \boldsymbol{\Lambda}_h^{-1}\right)\right) \\ &= -\mathrm{diag}(\bar{\boldsymbol{\Lambda}}^{-1})^{\mathrm{T}}\begin{bmatrix} \boldsymbol{\Lambda}_h^{-1} \\ \vdots \\ \boldsymbol{\Lambda}_h^{-1}\end{bmatrix} \end{aligned} \tag{92}$$

and

$$\begin{aligned} B_2 &= \frac{1}{M_s}\sum_{m=1}^{M_s} \bar{\mathcal{Y}}_m^H \mathbf{C}_{\bar{y}}^{-1}\left(\mathbf{I}\otimes \boldsymbol{\Lambda}_h^{-1}\right)\mathbf{C}_{\bar{y}}^{-1}\bar{\mathcal{Y}}_m \\ &= \frac{1}{M_s}\sum_{m=1}^{M_s} \bar{\mathcal{Y}}_m^H(\mathbf{V}_w \otimes \mathbf{I}_{M_rM_t})\bar{\boldsymbol{\Lambda}}^{-1} \\ &\quad \times (\mathbf{V}_w^H \otimes \mathbf{I}_{M_rM_t})\left(\mathbf{I}\otimes \boldsymbol{\Lambda}_h^{-1}\right) \\ &\quad \times (\mathbf{V}_w \otimes \mathbf{I}_{M_rM_t})\bar{\boldsymbol{\Lambda}}^{-1}\left(\mathbf{V}_w^H \otimes \mathbf{I}_{M_rM_t}\right)\bar{\mathcal{Y}}_m \\ &= \frac{1}{M_s}\sum_{m=1}^{M_s} (\bar{\boldsymbol{\Lambda}}^{-1}\bar{\mathcal{X}}_m)^H\left(\mathbf{I}\otimes \boldsymbol{\Lambda}_h^{-1}\right)\bar{\boldsymbol{\Lambda}}^{-1}\bar{\mathcal{X}}_m \\ &= \frac{1}{M_s}\sum_{m=1}^{M_s} (\bar{\boldsymbol{\Lambda}}^{-1}\mathcal{X}_m)^H\left(\begin{bmatrix} \mathrm{diag}\left(\boldsymbol{\Lambda}_h^{-1}\right) \\ \vdots \\ \mathrm{diag}\left(\boldsymbol{\Lambda}_h^{-1}\right)\end{bmatrix}\circ(\bar{\boldsymbol{\Lambda}}^{-1}\mathcal{X}_m)\right). \end{aligned} \tag{93}$$

Hence, (73) can be implemented efficiently as

$$\frac{\partial L(\bar{\mathcal{Y}})}{\partial \sigma_n^2} = A_2 + B_2. \tag{94}$$

Following similar steps as those used in deriving the efficient implementation of (67), it is straightforward to see that (80) can be implemented as

$$\begin{aligned} \frac{\partial L(\tilde{\mathcal{Y}})}{\partial(\boldsymbol{\theta}_h)_i} &= -\mathrm{diag}(\tilde{\boldsymbol{\Lambda}}^{-1})^H\begin{bmatrix} \mathrm{diag}\left(\mathbf{V}_h^H\frac{\partial \mathbf{C}_h}{\partial(\boldsymbol{\theta}_h)_i}\mathbf{V}_h\right) \\ \vdots \\ \mathrm{diag}\left(\mathbf{V}_h^H\frac{\partial \mathbf{C}_h}{\partial(\boldsymbol{\theta}_h)_i}\mathbf{V}_h\right)\end{bmatrix} \\ &\quad + \frac{1}{M_s}\sum_{m=1}^{M_s}(\tilde{\boldsymbol{\Lambda}}^{-1}\tilde{\mathcal{X}}_m)^H \\ &\quad \times \mathrm{vec}\left(\mathbf{V}_h^H\frac{\partial \mathbf{C}_h}{\partial(\boldsymbol{\theta}_h)_i}\mathbf{V}_h\mathrm{mat}(\tilde{\boldsymbol{\Lambda}}^{-1}\tilde{\mathcal{X}}_m, M_rM_t, M_f)\right) \end{aligned} \tag{95}$$

where the definition of $\tilde{\mathcal{X}}_m$ in (52) and the corresponding efficient implementation (53) were used.

The definition of matrices $\mathbf{D}_{1,wn}$ and $\mathbf{D}_{2,wm}$, $M_o^2 \times N_w$, in (77) and (78) do not take into account that several of their entries are equal to zero as a consequence of the Kronecker product in $(\partial \mathbf{C}_y)/(\partial(\boldsymbol{\theta}_w)_i)$. We define

$$\begin{aligned} \mathbf{A}_3(i) &= \mathrm{vec}\left\{\mathbf{C}_y^{-1}\left(\frac{\partial \mathbf{C}_w}{\partial(\boldsymbol{\theta}_w)_i}\otimes \mathbf{I}\right)\right\} \\ &= \mathrm{vec}\left\{\bar{\boldsymbol{\Lambda}}^{-1}\left(\mathbf{V}_w^H\frac{\partial \mathbf{C}_w}{\partial(\boldsymbol{\theta}_w)_i}\mathbf{V}_w \otimes \mathbf{I}\right)\right\}. \end{aligned} \tag{96}$$

We can define the $M_o \times 1$ vector $\mathbf{A}_3'$ containing the nonzero elements of $\mathbf{A}_3$ as

$$\mathbf{A}_3'(i) = \mathrm{diag}(\bar{\boldsymbol{\Lambda}}^{-1})\circ\begin{bmatrix} \mathcal{T}\left(\frac{\partial \mathbf{v}(\boldsymbol{\theta}_{wn})}{\partial(\boldsymbol{\theta}_w)_i}\right) \\ \vdots \\ \mathcal{T}\left(\frac{\partial \mathbf{v}(\boldsymbol{\theta}_{wn})}{\partial(\boldsymbol{\theta}_w)_i}\right)\end{bmatrix}. \tag{97}$$

Similarly, we can define

$$\begin{aligned} \mathbf{A}_4 &= \mathrm{vec}\left\{\mathbf{C}_{\bar{y}}^{-1}\left(\mathbf{I}\otimes \boldsymbol{\Lambda}_h^{-1}\right)\right\} \\ &= \mathrm{vec}\left\{\bar{\boldsymbol{\Lambda}}^{-1}\left(\mathbf{I}\otimes \boldsymbol{\Lambda}_h^{-1}\right)\right\}. \end{aligned} \tag{98}$$

$$\mathbf{D}_{1,wn} = \left[\operatorname{diag}(\bar{\mathbf{\Lambda}}^{-1}) \circ \operatorname{vec}\left\{ \begin{bmatrix} \mathcal{T}\left(\frac{\partial \mathbf{v}(\boldsymbol{\theta}_{wn})}{\partial(\boldsymbol{\theta}_w)_1}\right)^{\mathrm{T}} \\ \vdots \\ \mathcal{T}\left(\frac{\partial \mathbf{v}(\boldsymbol{\theta}_{wn})}{\partial(\boldsymbol{\theta}_w)_1}\right)^{\mathrm{T}} \end{bmatrix} \right\} \quad \cdots \quad \operatorname{diag}(\bar{\mathbf{\Lambda}}^{-1}) \circ \operatorname{vec}\left\{ \begin{bmatrix} \mathcal{T}\left(\frac{\partial \mathbf{v}(\boldsymbol{\theta}_{wn})}{\partial(\boldsymbol{\theta}_w)_{N_w}}\right)^{\mathrm{T}} \\ \vdots \\ \mathcal{T}\left(\frac{\partial \mathbf{v}(\boldsymbol{\theta}_{wn})}{\partial(\boldsymbol{\theta}_w)_{N_w}}\right)^{\mathrm{T}} \end{bmatrix} \right\} \right.$$
$$\left. \operatorname{diag}(\bar{\mathbf{\Lambda}}^{-1}) \circ \operatorname{vec}\left\{ \begin{bmatrix} \operatorname{diag}(\bar{\mathbf{\Lambda}}_h^{-1})^{\mathrm{T}} \\ \vdots \\ \operatorname{diag}(\bar{\mathbf{\Lambda}}_h^{-1})^{\mathrm{T}} \end{bmatrix} \right\} \right] \tag{101}$$

$$\mathbf{D}_{1,h} = \left[\operatorname{diag}(\bar{\mathbf{\Lambda}}^{-1}) \circ \operatorname{vec}\left\{ \begin{bmatrix} \left(\mathbf{V}_h^H \frac{\partial \mathbf{C}_h}{\partial(\boldsymbol{\theta}_h)_1} \mathbf{V}_h\right)^{\mathrm{T}} \\ \vdots \\ \left(\mathbf{V}_h^H \frac{\partial \mathbf{C}_h}{\partial(\boldsymbol{\theta}_h)_1} \mathbf{V}_h\right)^{\mathrm{T}} \end{bmatrix} \right\} \quad \cdots \quad \operatorname{diag}(\bar{\mathbf{\Lambda}}^{-1}) \circ \operatorname{vec}\left\{ \begin{bmatrix} \left(\mathbf{V}_h^H \frac{\partial \mathbf{C}_h}{\partial(\boldsymbol{\theta}_h)_{N_h}} \mathbf{V}_h\right)^{\mathrm{T}} \\ \vdots \\ \left(\mathbf{V}_h^H \frac{\partial \mathbf{C}_h}{\partial(\boldsymbol{\theta}_h)_{N_h}} \mathbf{V}_h\right)^{\mathrm{T}} \end{bmatrix} \right\} \right] \tag{102}$$

and

$$\mathbf{D}_{2,h} = \left[\operatorname{diag}(\bar{\mathbf{\Lambda}}^{-1}) \circ \operatorname{vec}\left\{ \begin{bmatrix} \mathbf{V}_h^H \frac{\partial \mathbf{C}_h}{\partial(\boldsymbol{\theta}_h)_1} \mathbf{V}_h \\ \vdots \\ \mathbf{V}_h^H \frac{\partial \mathbf{C}_h}{\partial(\boldsymbol{\theta}_h)_1} \mathbf{V}_h \end{bmatrix} \right\} \quad \cdots \quad \operatorname{diag}(\bar{\mathbf{\Lambda}}^{-1}) \circ \operatorname{vec}\left\{ \begin{bmatrix} \mathbf{V}_h^H \frac{\partial \mathbf{C}_h}{\partial(\boldsymbol{\theta}_h)_{N_h}} \mathbf{V}_h \\ \vdots \\ \mathbf{V}_h^H \frac{\partial \mathbf{C}_h}{\partial(\boldsymbol{\theta}_h)_{N_h}} \mathbf{V}_h \end{bmatrix} \right\} \right]. \tag{103}$$

We can define the $M_o \times 1$ vector $\mathbf{A}_4'$ containing the nonzero elements of $\mathbf{A}_4$ as

$$\mathbf{A}_4' = \operatorname{diag}(\bar{\mathbf{\Lambda}}^{-1}) \circ \begin{bmatrix} \operatorname{diag}\left(\bar{\mathbf{\Lambda}}_h^{-1}\right) \\ \vdots \\ \operatorname{diag}\left(\bar{\mathbf{\Lambda}}_h^{-1}\right) \end{bmatrix}. \tag{99}$$

Hence, in order to be implemented efficiently, $\mathbf{D}_{2,wn}$ can be redefined as

$$\mathbf{D}_{2,wn} = [A_3'(1) \quad \cdots \quad A_3'(N_w) \quad A_4']. \tag{100}$$

It is straightforward to see that $\mathbf{D}_{l,wn}$ can also be implemented efficiently as (101), shown at the top of the page, where the operator $\mathcal{T}(\cdot)$ is defined in (43).

Also, following similar derivations, we can implement (85) and (86) as (102) and (103), shown at the top of the page. Hence, the approximation of the Hessian can be expressed in a compact form as

$$E\left[\frac{\partial^2 L(\tilde{\mathcal{Y}})}{\partial(\boldsymbol{\theta}_h)_i \partial(\boldsymbol{\theta}_h)_k}\right] = M_s \mathbf{D}_{1,h}^{\mathrm{T}} \mathbf{D}_{2,h}. \tag{104}$$

REFERENCES

[1] M. Steinbauer, A. F. Molisch, and E. Bonek, "The double-directional mobile radio channel," *IEEE Antennas Propag. Mag.*, vol. 43, no. 4, pp. 51–63, Aug. 2001.

[2] B. H. Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus, and K. I. Pedersen, "Channel parameter estimation in mobile radio environments using the SAGE algorithm," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 3, pp. 434–450, Mar. 1999.

[3] A. Richter, "Estimation of radio channel parameters: Models and algorithms," Ph.D. dissertation, Elect. Eng. and Inf. Technol. Dept., Technische Universtät Ilmenau, Ilmenau, Germany, 2005, urn:nbn:de:gbv:ilm1-2005000111.

[4] J. W. Wallace and M. Jensen, "Modeling the indoor MIMO wireless channel," *IEEE Trans. Antennas Propag.*, vol. 50, no. 5, pp. 591–599, May 2002.

[5] A. Abdi and M. Kaveh, "A space-time correlation model for multielement antenna systems in mobile fading channels," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 550–560, Apr. 2002.

[6] A. Richter, M. Landmann, and R. S. Thomä, "Parameter estimation results of specular and dense multipath components in micro- and macrocell scenarios," presented at the Int. Symp. Wireless Personal Multimedia Communications (WPMC), Abano Terme, Italy, Sep. 2004.

[7] T. Trump and B. Ottersten, "Estimation of nominal direction of arrival and angular spread using an array of sensors," *Signal Process.*, vol. 50, no. 1–2, pp. 57–69, Apr. 1996.

[8] D. Astély, B. Ottersten, and A. L. Swindlehurst, "A generalized array manifold model for local scattering in wireless communications," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 1997, vol. 5, pp. 4021–4024.

[9] D. Astély and B. Ottersten, "The effects of local scattering on direction of arrival estimation with music," *IEEE Trans. Signal Process.*, vol. 47, no. 12, pp. 3220–3234, Dec. 1999.

[10] M. Bengtsson and B. Ottersten, "Low-complexity estimators for distributed sources," *IEEE Trans. Signal Process.*, vol. 48, no. 8, pp. 2185–2194, Aug. 2000.

[11] O. Besson and P. Stoica, "Decoupled estimation of DOA and angular spread for a spatially distributed source," *IEEE Trans. Signal Process.*, vol. 48, no. 7, pp. 1872–1882, Jul. 2000.

[12] Q. Wan, J. Yuan, and Y. N. Peng, "Estimation of nominal direction of arrival and angular spread using the determinant of the data matrix," in *Proc. 4th Int. Workshop Mobile Wireless Communications Network*, 2002, pp. 76–79.

[13] C. B. Ribeiro, E. Ollila, and V. Koivunen, "Propagation parameter estimation in MIMO systems using mixture of angular distributions model," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP 2005)*, Mar. 2005, vol. 4, pp. 885–888.

[14] K. V. Mardia, *Statistics of Directional Data*. New York: Academic, 1972.

[15] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall International, 1993.

[16] B. H. Fleury, P. Jourdan, and A. Stucki, "High-resolution channel parameter estimation for MIMO applications using the SAGE algorithm," *Proc. 2002 Int. Zurich Seminar BroadBand Communications. Access, Transmission, Networking*, pp. 30-1–30-9, Feb. 2002.

[17] R. S. Thomä, M. Landmann, A. Richter, and U. Trautwein, "Part II. Channel," in *Multidimensional High-Resolution Channel Sounding Measurement*. New York: Hindawi Publishing Corp., 2005, pp. 241–270.
[18] C. B. Ribeiro, E. Ollila, and V. Koivunen, "Stochastic maximum-likelihood method for MIMO propagation parameter estimation," *IEEE Trans. Signal Process.*, vol. 55, no. 1, pp. 46–55, Jan. 2007.
[19] S. Feng-Wen, J. Yimin, and J. Baras, "On the convergence of the inverses of Toeplitz matrices and its applications," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 180–190, Jan. 2003.
[20] P. Sherman, "Circulant approximations of the inverses of Toeplitz matrices and related quantities with applications to stationary random processes," *IEEE Trans. Signal Process.*, vol. 33, no. 6, pp. 1630–1632, Dec. 1985.
[21] D. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM J. Appl. Math.*, vol. 11, pp. 431–441, 1963.
[22] F. Belloni, A. Richter, and V. Koivunen, "Extension of root-music to non-ULA array configurations," presented at the IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), Toulouse, France, May 2006.
[23] G. H. Golub and C. F. V. Loan, *Matrix Computations*. Baltimore, MD: The Johns Hopkins Univ. Press, 1989.

Cássio Ribeiro (S'00–M'00) was born in Brazil in 1977. He received the Electronics Eng. degree from the Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil, in 2000 and the M.Sc. degree in electrical engineering from COPPE/UFRJ in 2002. He is currently working towards the D.Sc. degree at UFRJ and the Helsinki University of Technology, Helsinki, Finland.

In 2006, he worked as a Researcher at Nokia Institute of Technology, Brazil. Since 2007, he has been working at Nokia Research Center, Helsinki, Finland. His current research interests are in statistical signal processing, parameter estimation methods, and cellular systems.

Mr. Ribeiro coauthored a paper receiving the Best Paper Award in IEEE Personal, Indoor and Mobile Radio Communications (PIMRC) 2005.

Andreas Richter (M'04) was born in Germany in 1969. He received the Dipl.-Ing. (M.Sc.) degree in electrical engineering and the Dr.-Ing. (Ph.D.) degree (*summa cum laude*) from the Technische Universität Ilmenau, Ilmenau, Germany, in 1995 and 2005, respectively.

From 1995 to 2005, he worked as an Research Assistant at the Electronic Measurement Laboratory of Technische Universität Ilmenau. From July to October 2001, he was a Guest Researcher at the Wireless Laboratories, NTT DoCoMo, Yokosuka, Japan. Since 2004, he has been working as a Senior Researcher in the Statistical Signal Processing Laboratory at Helsinki University of Technology, Finland. His research interests are in the fields of digital communication, sensor-array, and statistical signal processing.

Dr. Richter was coauthor or author of five papers receiving a Best Paper Award (EPMCC'01, ISAP'04, PIMRC'05, EUSIPCO'06, and EuCAP'06). In 2005, he received the Siemens Communications Academic Award. For their work on MIMO channel sounding, he and his former colleagues at Technische Universität Ilmenau received the Thuringien Research Award for Applied Research in 2007.

Visa Koivunen (S'87–M'87–SM'98) received the D.Sc. (Tech.) degree with honors from the Department of Electrical Engineering, University of Oulu, Finland.

From 1992 to 1995, he was a visiting researcher at the University of Pennsylvania, Philadelphia. From 1997 to 1999, he was an Associate Professor at the Signal Processing Laboratory, Tampere University of Technology, Tampere, Finland. Since 1999, he has been a Professor of signal processing at the Department of Electrical and Communications Engineering, Helsinki University of Technology (HUT), Finland. He is one of the Principal Investigators in the SMARAD (Smart and Novel Radios) Center of Excellence in Radio and Communications Engineering nominated by the Academy of Finland. Since 2003, he has been Adjunct Professor at the University of Pennsylvania. During his sabbatical term 2006–2007, he is Nokia Visiting Fellow at Nokia Research Center, Helsinki, and visiting fellow at Princeton University, Princeton, NJ. His research interests include statistical, communications, and sensor array signal processing. He has published more than 200 peer-reviewed papers in international scientific conferences and journals.

Dr. Koivunen coauthored papers receiving the Best Paper Awards in IEEE PIMRC 2005, EUSIPCO 2006, and EuCAP 2006 conferences. He served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS. He is a member of the editorial board for the *Signal Processing* journal and the *Journal of Wireless Communication and Networking*. He is also a member of the IEEE Signal Processing for Communication Technical Committee (SPCOM-TC). He is the General Chair of the IEEE Signal Processing Advances in Wireless Communication (SPAWC) 2007 conference in Helsinki, June 2007.